CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering
Department of Radioelectronics

Quality of Audiovisual Signals

Master Thesis

Study Programme: Communication, Multimedia, and Electronics
Branch of study:Multimedia Technology

Thesis advisor:Libor Husnik

Yevgeniya Tyumina (author)

Prague 2014

Czech Technical University in Prague
Faculty of Electrical Engineering

Department of Radioelectronics

DIPLOMA THESIS ASSIGNMENT

Student: **Yevgeniya Tyumina**

Study programme: Communications, Multimedia, Electronics
Specialisation: Multimedia Technology

Title of Diploma Thesis: **Quality of Audio - Visual Signals**

Guidelines:

Prepare the subjective test with the aim to study mutual interaction of quality of individual components of audio-video signal on the overall quality. Prepare the testing material, test scheme and perform the test with at least 20 subjects. For analysis of test results use ANOVA statistics.

Bibliography/Sources:

[1] Guilford,J.: Psychometric Methods, McGraw Hill, 1954
[2] Herrera, M.: Evaluation of Audio Coding Artifacts, PhD Thesis, CTU FEE 2009
[3] ITU-R Recommendation BS.1116-1

Diploma Thesis Supervisor: Dr.Ing. Libor Husník

Valid until the end of the summer semester of academic year 2014/2015

Prof. Ing. Miloš Klíma, CSc.
Head of Department

L.S.

prof. Ing. Pavel Ripka, CSc.
Dean

Prague, February 10, 2014

Abstract

The aim of the presented Master Thesis is to find a mutual interaction between audio quality, video quality and contents of audiovisual sequences and conclude how much these three parameters influence on overall audiovisual perception.

A practical tool which was applied in the presented work was a subjective quality measurement. In the theoretical part there are described several possible psychometrics methods which can be used for the testing and also the principle of T-test and ANOVA analyses.

The test was prepared based on the method of successive categories using ACR-5. As stimuli base there were recorded 5 different content audiovisual sequences. For the test there were used audiovisual signals with combinations of 5 levels of audio and 7 levels of video qualities.

The data from the measurement were analysed using T-test and ANOVA analyses.

Acknowledgement

First and foremost I would like to thank professor **Miloš Klima** for his guidance and provision of information in initial steps of the research. My great gratitude is to my thesis supervisor **Libor Husnik**. It would be almost impossible to accomplish the work without his assistance and dedicated involvement. Also I would like to thank **Karel Fliegel** for professional consultation during the preparation of audiovisual sequences for the test.

It is impossible not to mention in my gratitude **Jaroslav Bouše** who accompanied and helped me during the practical part of the research. I would like to thank **Jan Bednář** for providing me with subjects for the test and the other **colleagues from Ragio Engineering department** for their participation in the measurement and overall support.

Special thanks to my friend **Aigera Karimova**, whose "amazing time" made sufficient contribute to science.

I would like to thank my **dear family and friends** whose intensive support I have been gotten during all my studies from different countries through thousands of kilometres.

And also I want to thank **Robert McFerrin** for his song "Don't worry, be happy" which helped me to overcome all troubles and difficulties during the work.

Author

NOTATION

σ_1^2 - group of variances

ACR-5 - Absolute Category Rating 5 grade scale

ACR-9 Absolute Category Rating 11 grade scale

ANOVA analysis of variance

$C_2(n)$- number of pairs regarding the order

d - the number of triads in a given measurement

DBTS Double blind triple stimulus with hidden references method

df_A – degrees of freedom between groups

df_{AV}-degrees of freedom of interaction between groups

df_{total} - Total degrees of freedom

df_w - Degrees of freedom (within groups)

Fps frame per second

F-ratio – variance ratio

H_0 - zero hypothesis

H_A – alternative hypothesis

HD high definition

K - Kendal's consistency coefficient

k - number of judges

k_A - number of groups in the condition of the first factor

k_V - number of groups in the condition of the second factor

MS_{total}- Total mean of squares

MS_w-Mean square (within group)

N - total number of all elements in the analysis

n - total number of observers

n_1 - sample size

n_w - number of elements within one group

S - sum of ranks for any stimulus.

SS_A , SS_V – sums of squares between groups

SS_{AV} – sums of squares between factors

SS_{total} – total sums of squares

SS_w- sums of squares within groups

$V_2(n)$ - The number of pairs regardless the order

x_{si}- mean value of evaluated stimulus i in s-th trial

x_s- mean value of all stimuli in s-th trial

x_i - evaluation from stimuli i,

s_{si}- standard deviation of all stimuli in in s-th trial

s_s - standard deviation of evaluated stimulus i in s-th trial

$\bar{p}$ - average intercorrelation among individual judges

d_{max} - the maximum number of circular triads

X - value of observations

List of figures

List of tables

Contents

Introduction

The history of video transmission has started from the second part of XIX century, when in 1873 a German inventor Paul Nipkow discovered a so called Nipkow's disk [1]. The disk had 18 squared apertures arrayed into a spiral pattern around it. Nowadays his invention seems rather primitive, but those days it was a great scientific breakthrough. From that time a progress in video industry has never stopped. Almost for two centuries the evolution has developed from numb black and white signals to high resolution signals, containing full coloured pictures and natural sounded audio tracks. The possibility of video cameras in addition to special effects allows to directors of movies create impossible. The only thing which limits actual video contents is imagination. But the more digital information contains a video signal the more bitrates it requires. The more bitrate has a video, the more bandwidth it requires for transmission. It is fair as for television transmission as for Internet one. The urgent requirements for videos are the highest possible quality with the lowest possible bitrate.

The main idea of the theme is to find interaction between three parameters which influence on overall audiovisual perception. The parameters are contents of sequences, audio and video quality. It is presumed that depending on the first parameter the influence of the second and the third parameters are variable.

The theme of evaluating overall audiovisual quality is not new and some research has already been done in this field. For example in December 2004 David S. Hands published an article about dependence of overall video quality on content of the video. The results showed that for stable video such as "head and shoulders" the quality for audio should be higher than for video, and for high motion videos the quality should be vice versa [2]. Julie Lassalle in a work "Impact of the content on subjective evaluation of audiovisual quality"[3] proved that depending on a video content different distortions are perceived less or more annoying. In a testing material were presented such distortions as audio delay, frozen frame, packet losses. She also concluded that different content videos demand variable audio and video quality. Also were made some tries to optimise transmission of video by estimating speech/non-speech parts of audio stream based on the same idea [4]. Stefan Winkler and Christof Faller presented a work with different content videos where they estimated audio and video quality separately and together. As a result they showed that the given bitrate should be distributed

between audio and video not uniformly [5]. There is also some works dedicated to optimal test materials content where described some parameters which should be presented in stimuli to get maximum objective results. For example Woszczyk in his work [6] suggested a special set of perceptual dimensions of audio-visual experience such as space, motion, mood and action, which observer needs to perceive when watching videos. And in the same time this dimensions are characterised by four attributes: quality, magnitude, involvement and balance. And later Ulrich Reiter in his work "Overall perceiver audiovisual quality – what people pay attention to" [7] reduced the 4x4 set to more compact design of 3x2 set (dimensions: action, mood, space and attributes of averaged value of quality-magnitude-involvement and balance) which can provide the same results without losing perceived information.

The presented Master thesis consists of three chapters conditionally divided into three parts: theoretical, practical and analytical. The first chapter is dedicated to theoretical description and analysis of possible psychometrics methods for the research. The second part, practical one, contains information about general recommendations for audiovisual tests and also full description of actual test for the research. The description of the test contains particular information about audiovisual sequences used as stimuli and also specification of test conditions. The third chapter deals with theoretical principles of a method of statistical analysis which is going to be applied for interpretation of test data. In the last chapter there is presented actual analysis of the research with the following conclusion.

1 Analysis of the task

According to the theme of the current diploma thesis the main idea of the project is to establish dependence between overall audiovisual qualities by influence of contents of audiovisual sequences.

To obtain the proper results the following steps must be done:

1. Analysis of previous research on resembling theme.
2. Studying of available methods of testing and its analysis
3. Design of the test structure
4. Preparation of stimuli for the test
5. Preparation of the test conditions
6. Pre-test few subjects before the actual test to be sure that conditions are appropriate
7. Test subjects
8. Analysis of the test data
9. Conclusion of the work.

The first two steps are described in the Chapter 2 and 4; the next 5 steps are described in the fourth chapter. The analysis is presented in the Chapter 5 and the results are concluded in the last chapter.

2 Review of psychometric methods

The most essential part of the project is testing. That is why a method which is used for that is very important because it defines the direction of the work. For this reason here will be introduced several methods which can be used for the testing and then will be made a conclusion which one is more preferable for this kind of work.

It is also important that observers'decision about overall quality depends on a lot of aspects such as mood in the moment of testing, irksome of sequences, content of the stimuli (Graham's equation). That is why to get more appropriate results in there should be used some psychometric methods. In the paragraphs below there are discussed methods of average error, minimal changes, constant, pair comparison, rank order method, and successive categories [8].

2.1 The method of average error

This method based on comparing compressed stimuli with a reference one. The observer is able to actively participate in variation of stimuli by manually controlling their properties to make them closer to the reference. Fig 2.1 presents the schematic variant of this method.

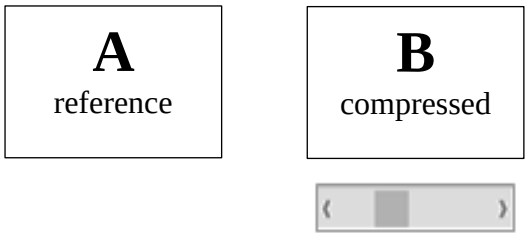


Fig 2.1 Schematic variant of the method of average error

This method is useful when a measurement must be done in limited time because every trail is able to give some particular result, when in other methods this result obtains by some sequences of trails. Such kind of method is suitable for evaluating images, but not for videos. For video it is not enough just at one "glance" to notice some difference and it is needed to be watched till the end.

2.2 The method of minimal changes

This method uses the principle of recognizing just noticeable difference. There is also presented a reference stimulus (as it can be seen in Fig.2.2) and with a decreased or increased quality of the other stimuli observer should report when he had perceived just noticeable difference. But in contrast to previous method of average error here an observer has no active participation in variation of stimuli properties.

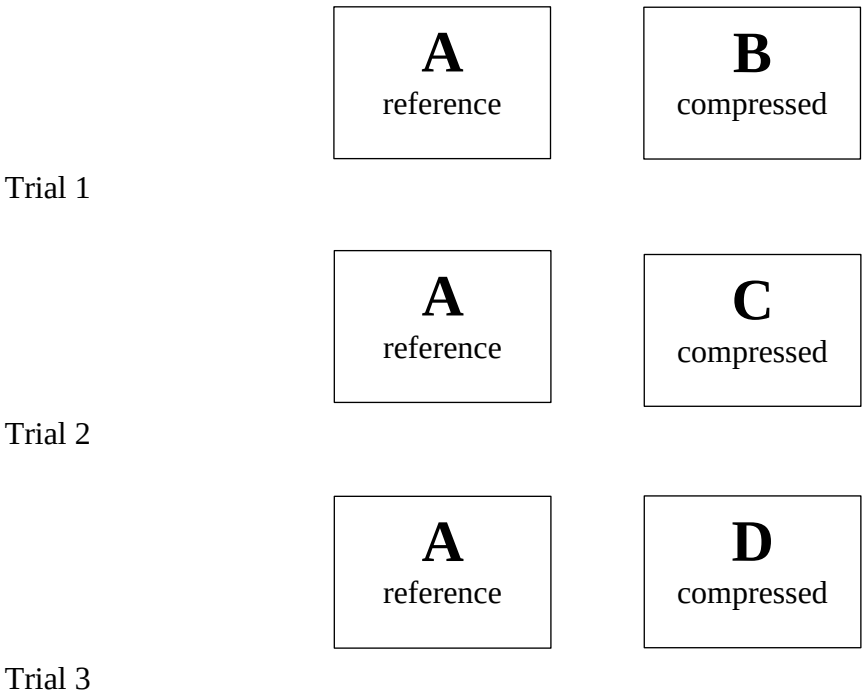


Fig 2.2 Schematic variant of the method of minimal changes. Quality of compressed stimuli is increased or decreased consequentially.

In this method can be faced some errors in perceiving such as habituation and expectation errors. The first one is connected with tiny difference of the stimuli and long series of the same stimuli. The second error occurs when after some trials an observer can get some expectation of changing quality even it is the same. That is why the size of a step between neighbour stimuli and sequence of presenting stimuli are the key moments in this measurement.

2.3 The method of double blind triple stimulus with hidden reference

The method of double blind triple stimulus with hidden reference (DBTS) refers to the method of minimal changes. In one trail there are presented two reference and compressed

stimuli where only one reference stimulus is titled another one is hidden. An example of DBTS method is shown in Fig 2.3.

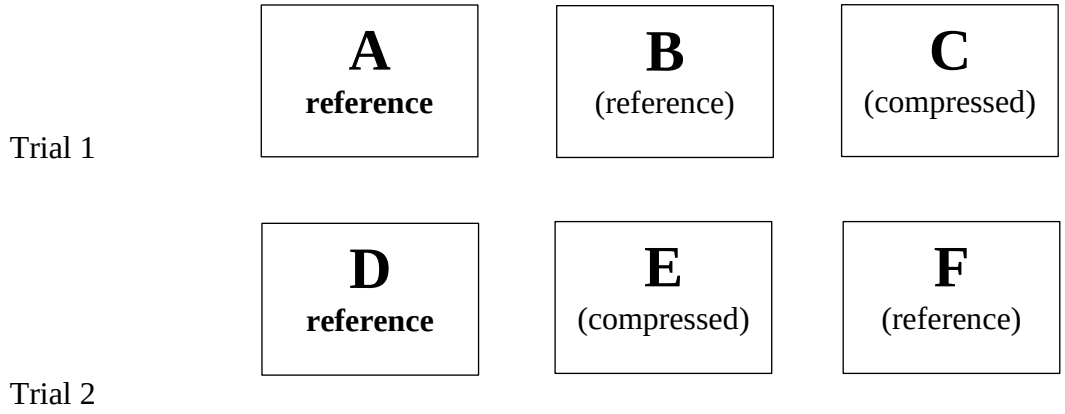


Fig 2.3 Schematic variant of the method of DBTS

The task of an observer is to find the reference (original) stimulus and also to evaluate the quality of compressed stimulus using evaluation scale (table 1).

Impairment	grade
Imperceptible	5
Perceptible, but not annoying	4
Slightly annoying	3
Annoying	2
Very annoying	1

Table 1 DBTS evaluation scale

Normalization of results Z_i is calculated by a formula 1.1:

$$Z_i = \frac{x_i - x_{si}}{s_{si}} \cdot s_s + x_s \qquad 1.1$$

Where x_i - an evaluation from stimuli i,

- x_{si}- a mean value of evaluated stimulus i in s-th trial
- x_s- a mean value of all *stimuli* in s-th trial
- s_{si}- a standard deviation of all stimuli in in s-th trial
- s_s - a standard deviation of evaluated stimulus i in s-th trial

An observer must know the test scheme. Due to tiredness the testing time should not be more than 30 minutes.

2.4 The constant method

The method is considered to be the most accurate of all the psychophysical methods. Compared to other methods it requires small number of stimuli (4-7) which are presented large number of times (50-200).

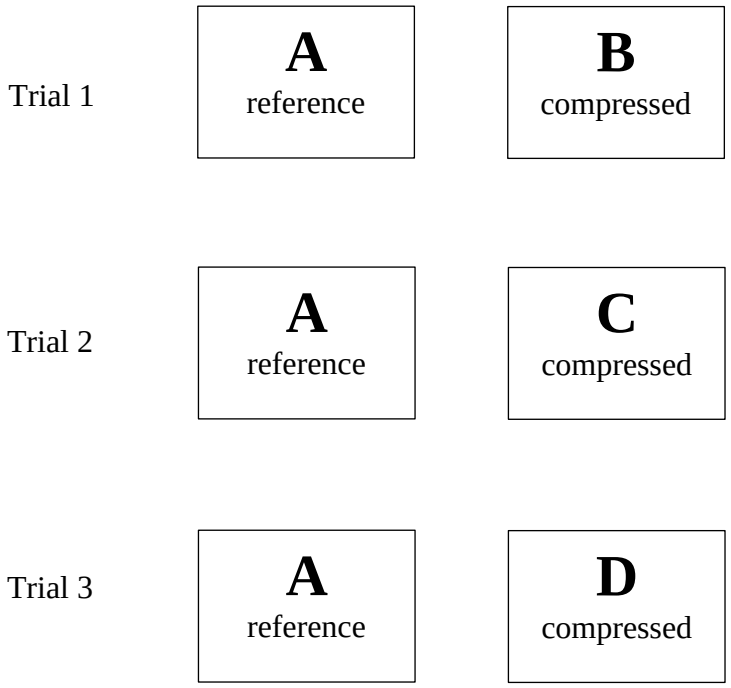


Fig 2.4 Schematic variant of constant method. Compressed stimuli are presented randomly.

In this method some perceived errors can be eliminated by repetition. But despite of advantages the constant method is not the best approach for the project case. In the project there must be present rather bigger number of stimuli than 4-7 like in this method. Otherwise the observer will be loaded to evaluate too many sequences of stimuli.

2.5 The method of pair comparison

In this method in every trial the observer needs to compare two stimuli where no reference stimulus is presented. An order of pair sequences must be determined in advance.

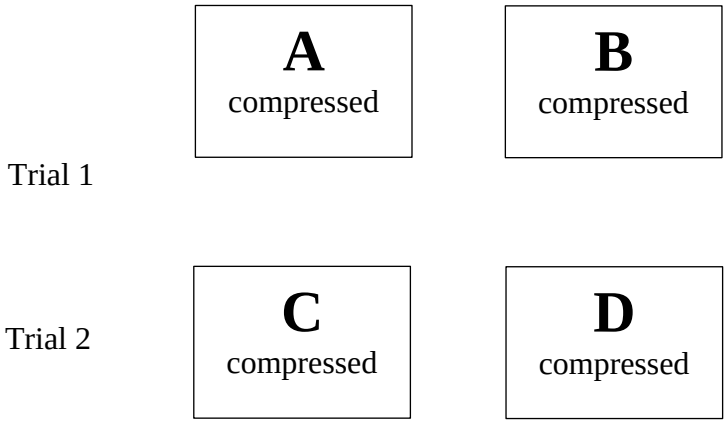


Fig 2.5 Schematic variant of the method of pair comparison

This method requires no physical scaling, so the same pairs stimuli can be evaluated by many various aspects. The core of the method is in comparison of two stimuli only by guessing which one is better. The simplicity for observer is referred to the fact that the difference in the attitude to the stimuli can be easily determined. The only critical point in limiting number of trails is in the composition of the pairs and their further sequences. One stimuli can either be presented in one pair once either in various pairs in different series. Also due to specificity of video evaluation this method is sensitive to all individual requirements of observer, can contain not that large number of pair sequences covering evaluation of overall properties.

The number of pairs regarding and regardless the order is determined by formulas 2.1 and 2.2 respectively.

$$C_2(n) = \frac{(n^2 - n)}{2} \tag{2.1}$$

$$V_2(n) = (n^2 - n) \tag{2.2}$$

where n - the number of stimuli

For pairs composition is used the rule of Latin Square: each stimulus can be used once at every position in a trail (once as the first one, once as the k-th one).

For arranging the order of pairs is used Ross's plan, where all stimuli are alternated and the same stimulus can be shown again in different pair at least in one trail. An example of Ross's plan arrangement is shown on a table 2.

1-2	3-5	4-1	2-3	5-4	1-3	4-2	5-1	3-4	2-5

Table 2 Ross's plan order

At the Ross's plan stimuli in pairs should be alternated along with distance between the same stimuli in different pairs at least one trail.

To check consistency of individuals' judgements is used Kendal's consistency coefficient (2.3).

$$K = 1 - \frac{d}{d_{max}} \tag{2.3}$$

-where *d* is the number of triads in a given measurement,

- d_{max} is the maximum number of circular triads.

The circular triads are formed as three sets of inequality: if A<B and B<C, then A<C.

For odd number of stimuli the coefficient is:

$$K = 1 - \frac{24d}{n^2 - n} \tag{2.4}$$

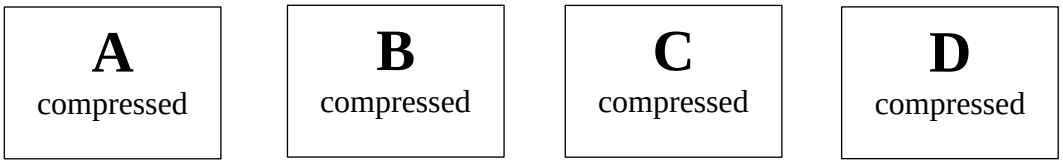
For even stimuli number Kendal's coefficient is:

$$K = 1 - \frac{24d}{n^2 - 4n} \tag{2.5}$$

2.6 The method of rank order

In this method stimuli can be judged with the reference of one to another. The observer is able to see at one moment several stimuli and then set the priority order. This method requires less time because of simultaneous evaluation of several stimuli.

The method of rank order is assumed to be the most popular and practical one but again it doesn't fit to all requirements of the project, because in this case it is quite problematic to focus on some proper stimulus. In the project both audio and video quality must be evaluated simultaneously. The overall quality depends on many aspects such as content of the video and evaluation of it should me more sensitive.



Order number	Stimuli conformity
1	
2	
3	
4	

Fig 2.6 Schematic variant of the method of rank order

2.7 The method of successive categories

The method of successive categories is a rating method with some predefined rating scale. The method is one of the most general methods of data evaluation but still is wide used. The rating scale is continuous and includes a limited number of categories. This method is a multidimensional method due to evaluation or comparison of different parameters during the test. For example in the case of the project the same video sequence can be evaluated by only perceived audio quality, video, or overall quality, and the results can be completely different. Below is depicted a schematic view of the method and tables 3 and 4 provide with some examples of the rating scales[9],[10].

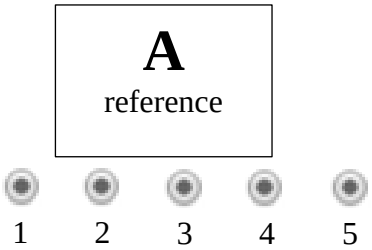


Fig 2.7 Schematic variant of the method of successive categories for single stimulus

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 3 Rating scale, for single stimulus method, example 1

5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table 4 Rating scale for single stimulus method, example 2

The example which is shown in Fig 2.7 is a single stimulus rating method. And observer has to evaluate all stimuli one by one using the same condition with the same rating scale for each parameter. The method also can be used for comparison reference and compressed stimuli. An observer has to compare two stimuli and evaluate a difference between them using a rating table. An example of the table is presented in Table 5.

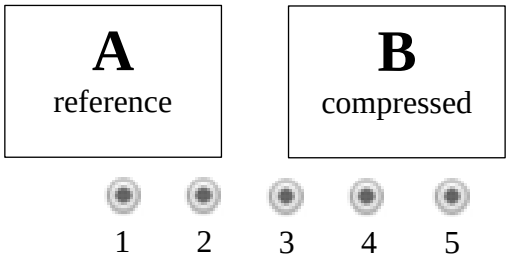


Fig 2.8 Schematic variant of the method of successive categories for comparison

3	Much better
2	Better
1	Slightly better
0	The same
-1	Slightly worse
-2	Worse
-3	Much worse

Table 5 Comparison rating scale

3 Preparation of the test

3.1 Recommendations for an audio-visual test

Source signal

Regarding recommendation on subjective audio-visual quality assessment [11] originals of audio-visual sequences must be recorded in the highest quality and a length of each of them should be around 10 s, but not less than 8 s. Also one of requirements is that sequences must have a logical ending without any interrupted phrase neither in speech neither in musical lyrics. There must be presented at least 4 different contents of audio-visual sequences in a trial set to avoid fatigue from monotony of the testing. To get more objective results from testing there must be excluded a personal attitude of observer to evaluated videos. To obtain that the selected videos should be with neutral expression, without any high-emotional details.

Conditions

Conditions for a particular test must be specified and fixed to make it equal for all subjects. Video sequence is recommended to be shown in a full-screen format. The size of a test room matters only when use amplifiers as an audio transmitted device.

Subjects

The number of subjects is fluctuating between 6 and 40, where 6 is an absolutely minimum, so it is recommended to have at least 15 subjects in the experiment. As the test subject can be accepted any age person with normal or corrected-to-normal visual and aural acuity [8],[9]. An observer should not be an experienced assessor and not have a previous direct involvement to picture or audio evaluation.

Instructions

Before the test starts all the subjects should be instructed with the test scenario and provided with description of the test and detailed rating scale in a written form. Also before the actual experiment subjects must be provided with at least 5 preliminary trials as examples of future test sequences to clarify the experimental task. Results from preliminary trials will be not included into experiment results.

3.2 Test set up

Source signal

For the experiment were recorded 5 different contented video sequences: 4 sequences with Prague views and one with a short monologue of Aigerim Karimova (Fig 3.1). Videos were captured with CANON 70D camera, full HD (1920/1080i, 25fps), sounds were recorded with 48 kHz sample frequency.

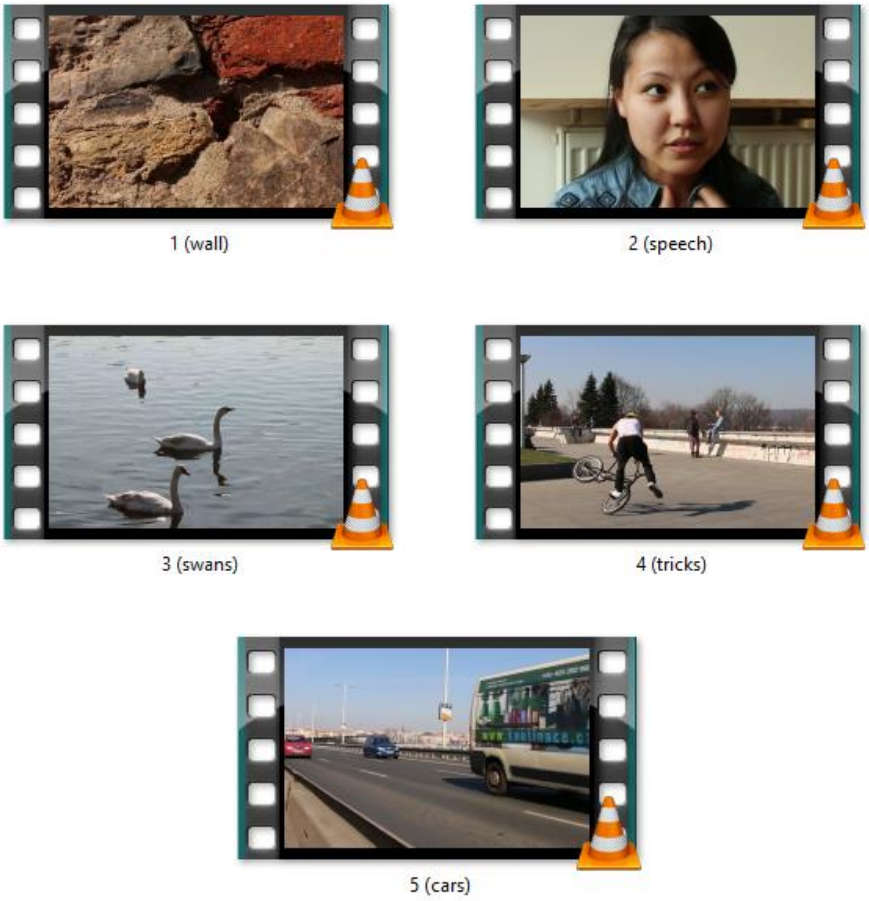


Fig 3.1 Test audio-visual sequences

For visual content were chosen scenes with variable emphasis on audio-visual perception. (The list with contends description is shown on a table 6). Thus videos with high motion require more visual attention then aural, and vice-versa head-and-shoulder sequences require more aural attention then visual.

Some audio tracks form original audio-visual sequences were replaced with exterior musical tracks (BirdPen “Off”, Ludovico Einaudi “Two Trees”, and Caribou “Odessa”) which are follow the style of the videos to make it more entertaining for overall perception.

№	name	description	length, sec
1	Wall	Slow motion camera movements, a lot of particularities on the picture, not original background music without lyrics	10
2	Speech	Head-and-shoulders, original sound	13
3	Swan	Big slow moving objects, background has no particularities, not original background music without lyrics	10
4	Tricks	Fast moving object, not original background music with lyrics	12
5	Cars	Stable background with big fast moving objects, original sound (ambient noise)	10

Table 6 – Audio-visual content description

The length of each video is about 10s depending on logical end of the scene. From each original video were produced 35 video sequences with 7 various video and 5 audio bitrates (varieties of bitrate are shown in tables 7,8). The derivative videos were compressed by H.264. Thus from 5 original videos were produced 175 new video sequences which going to be presented as test stimuli.

video, Mbps							original
0.5	1	2	5	10	30	70	80

Table 7 Variety of video qualities

audio, kbps					original
16	32	64	128	320	1536

Table 8 Variety of audio qualities

The experimental set contains 100 audio-visual sequences by 20 sequences per content. The order of sequences in an experimental set was chosen randomly, but in that way that a distance between two the same content videos is at least two positions. A list with sequences order is shown in Appendix 1.

In the beginning of the experiment there are an example set of five audio-visual sequences visually demonstrating the task and letting an observer to adjust volume and comfortable distance between a seat and a monitor.

Evaluating scale

For the experiment was used five-level quality scale. According to work of Japanese researches [12] were concluded, that five-grade absolute category rating scale (ACR-5) is more suitable for quality assessment tests due to its simplicity if compare it with more particular ACR-11 scale.

Grade	Equivalent
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Table 9 Evaluating scale

Conditions

Total, time of the experiment takes 28 minutes, that is why was taken decision to divide it into two parts by 14 minutes (48 and 52 audio-visual sequences for evaluation). Each part of experiment was exanimated in different days to prevent a fatigue and loss of concentration.

All the measurements were done in an acoustic booth which is situated at Czech Technical University at Radio Engineering department (Fig 3.2).

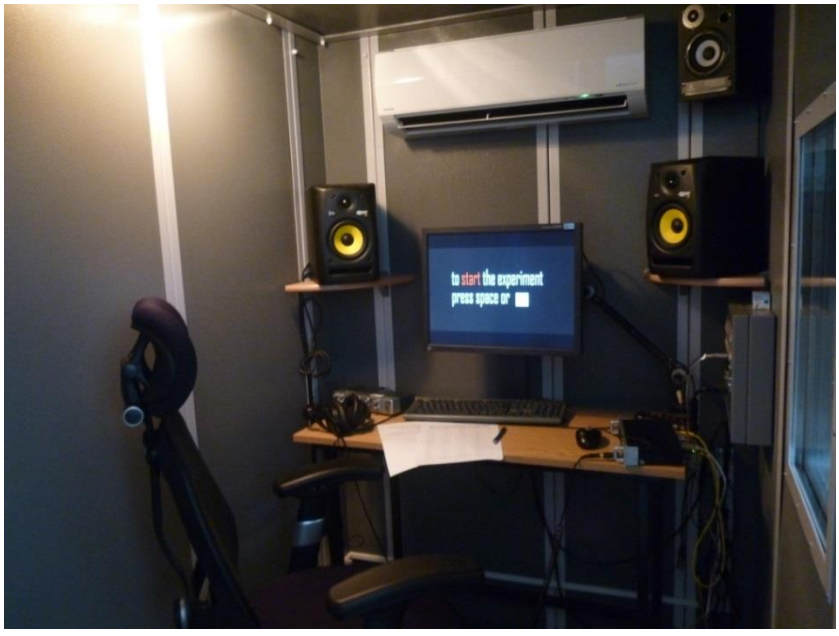


Fig 3.2 Booth for audiology measurement

The acoustic booth was mounted by the SONIG Company. The booth provided the experiment with the same conditions for each subject. It isolates the ambience noise and also excludes any possibility of interrupting during the experiment (the average energy spectrum of sound pressure level inside and outside the booth can be found in appendix 2[16])

The list of equipment which was used in an acoustic measurement booth during the experiment is listed below in a table 10.

Item	Name of a model
Headphones	Sennheiser HD 280 Pro
Soundcard	RME FIREFACE UC
Screen	EIZO Flex Scan S2000

Table 10 Equipment list

The screen in a measurement booth was measured by calibration probe (model i1pro). Average colour deviation delta 95 was equal to 0.25.

All the audio-visual sequences were played with a VLC Media Player in a full-screen mode.

Instructions

Before the experiment starts an observer was provided with oral instruction. Also at his/her disposal were written instructions with a short explanation and also an evaluating scale (Appendix 3), answer sheet (Appendix 4), and a pen.

Subjects

All the participated subject in the experiment were within an age range between 21 and 31 years with normal hearing and normal or corrected vision abilities. All subjects were students, but in different fields of study. 3 subjects studied film/photography, and had no experience at evaluation of any technical parameters neither audio, neither video. Other subjects were students of CVUT: 7 from transportation and mechanical engineering department (no experience with audio/video/audiovisual evaluation), and the rest from radio engineering department (some participated at audio or video evaluations, but non at audiovisual one). Thus all the subjects, participated in the testing, were not experienced with audiovisual test evaluation.

4 Methods of analysis

The next step after test completion is a data analysis. In this work there are going to be implemented two methods of data analysis: T-test and ANOVA. Implementation of two different methods for the same data allows ensuring more precise result.

4.1 T-test

T-Test is a method of statistical analysis aimed to evaluate difference between means of two groups of variances. For a proper T-test must be obeyed the following assumption:

- variables of each group must be independent
- each group is considered to be a sample from a distinct population
- all variables must have a normal distribution

In the beginning of the analysis there must be determined two hypothesis which will be accepted or rejected after its completion. The first hypothesis is a null hypothesis (H_0), which usually implies no significance difference between two groups of variances (4.1). The second hypothesis is an alternative one (H_A) which is an opposition of the first hypothesis (4.2).

H_0: $\mu_1=\mu_2$ (4.1)

H_A: $\mu_1 \neq \mu_2$ (4.2)

Where μ_1 and μ_2 are mean values of group 1 and group 2 correspondingly.

The corresponding calculation for T-test analysis is a calculation of t-value (4.3):

$$t - value = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (4.3)$$

Where σ_1^2 and σ_2^2 are group variances and n_1 and n_2 are sample size.

The acceptance or rejection of hypotheses depends on a significance level. The significance level is usually set as 5%, which is determines an appropriate reference distribution (t-distribution).

If the observed *t*-value is larger than the reference then the hypothesis is considered to be rejected.

4.2 Analysis of variance

Analysis of variance (ANOVA) is a statistical technique which is used to determine difference between variance inside a factor. An advantage of this technique is a possibility to work out data from more than two factors.

In ANOVA analysis four assumptions must be following:

- All the samples must be independent
- All groups must have the same number of samples
- The population of the samples must be normally distributed
- The variances of the populations must be equal

The analysis can be realized in three main steps: formation of hypotheses, filling a calculation table, taking decision which hypotheses is true.

4.2.1 Formation of hypotheses

Before starting with the first step, there must be clarified two criteria which are operated in ANOVA: factors and levels. As a factor we understand a changed parameter and as a level – variance of that parameter. There can be used one-factor, two, or multi-factor analysis. In a case of the project a two-factor analysis is used because there are only two changed parameters: audio and video quality; as levels are used 7 different bitrates of video and 5 different bitrates of audio, so in the analysis are two factors with 7 and levels correspondingly [13].

The first step after taking decision about factors and levels is a formation of hypotheses. In all cases of analysis there are two hypotheses: null and alternative one. The null hypothesis (H_0) contends that all samples in one factor have the same mean. The alternative hypothesis (H_1) contends the opposite: not all means are equal (at least one mean is different from other).

H_0: $\mu_1=\mu_2=\ldots=\mu_k$ (4.4)
H_1: some $\mu_k \neq \mu_k^*$ and H_0 is not true
Where μ_k is "any population mean" and μ_k^* is "anther population mean"

The only difference between hypotheses formation of one and two/multi factor analysis is that in the second case there are several sets of H_0 and H_1. According to the project task there are

going to be three sets of hypotheses. The first one (set A) defines relationships between means of video sequences with the constant video quality and variable audio quality, the second one (set V) – constant audio qualities and variable video qualities, and the third one (set AV) – an interaction between variable audio and video qualities.

Set A (4.5)

H_0: $\mu_{a1}=\mu_{a2}=\ldots=\mu_{ak}$

H_1: some $\mu_{ak}\neq\mu_{ak}$* and H_0 is not true

Set V (4.6)

H_0: $\mu_{v1}=\mu_{v2}=\ldots=\mu_{vk}$

H_1: some $\mu_{vk}\neq\mu_{vk}$* and H_0 is not true

Set AV (4.7)

H_0: $\mu_{av1}=\mu_{av2}=\ldots=\mu_{avk}$

H_1: some $\mu_{avk}\neq\mu_{avk}$* and H_0 is not true

Where μ_{ak} is "any population mean"

- μ_{ak}* is "another population mean"

indexes *a,v* and *av* are corresponding to variable audio, video and audiovisual quality.

Then it must be determined how accurate will be the decision about the hypotheses confirmation. Usually significance level is set up 95-99% and as following p=0.05 or p=0.01(Table 8, 9).

4.2.2 Calculation table

The calculation table contains data from all computational procedure of the analysis. Below is depicted an example of such a table according to the case of the project (Table 11)[14].

Source of variation	Degrees of freedom	Sums of squares	Mean square	Variance ratio (F)
Factor A	$\boldsymbol{df_A}$	$\boldsymbol{SS_A}$	$\boldsymbol{MS_A}$	$\boldsymbol{F_A}$
Factor V	$\boldsymbol{df_V}$	$\boldsymbol{SS_V}$	$\boldsymbol{MS_V}$	$\boldsymbol{F_B}$
Factor AV	$\boldsymbol{df_{AV}}$	$\boldsymbol{SS_{AV}}$	$\boldsymbol{MS_{AV}}$	$\boldsymbol{F_{AV}}$
Within groups	$\boldsymbol{df_w}$	$\boldsymbol{SS_w}$	$\boldsymbol{MS_w}$	
Total	$\boldsymbol{df_{total}}$	$\boldsymbol{SS_{total}}$	$\boldsymbol{MS_{total}}$	

Table 11 – Example of calculation table

Degrees of freedom (between groups)

Degrees of freedom between groups define as the value which is one less than the number of levels in a factor [15].

$$df_A=k_A-1 \quad (4.8)$$

$$df_V=k_V-1 \quad (4.9)$$

Where k_A number of groups in the condition of the first factor

- k_V number of groups in the condition of the second factor

The degree of freedom of interaction between two factors defines as a multiplication product of their degrees of freedom.

$$df_{AV}=(k_A-1)(k_V-1) \quad (4.10)$$

Total degrees of freedom

$$df_{total}=N-1 \quad (4.11)$$

where N is a total number of all elements in the analysis

Degrees of freedom (within groups)

$$df_w=k_A \cdot k_V(n-1) \quad (4.12)$$

Where n is a total number of observers

Sums of squares (between groups)

$$SS_A=\sum\left[\frac{(\sum X)^2}{n_V}\right]-\frac{(\sum X)^2}{N} \quad (4.13)$$

$$SS_V=\sum\left[\frac{(\sum X)^2}{n_A}\right]-\frac{(\sum X)^2}{N} \quad (4.14)$$

The sum of squares for interaction between two factors (SS_{AV}) can be found from the formula of the total sum of squares (3.13)

$$SS_{AV}=SS_{total}-SS_A-SS_V-SS_w \quad (4.15)$$

Total sums of squares

$$SS_{total}=SS_A+SS_V+SS_{AV}+SS_w \tag{4.16}$$

$$SS_{total}=\sum X^2-\frac{(\sum X)^2}{N} \tag{4.17}$$

Where X is a value of each observation

Sum of squares (within groups)

$$SS_w=\sum X^2-\sum\left[\frac{(\sum X)^2}{n_w}\right] \tag{4.18}$$

Where n_w is a number of elements within one group

Mean square (between groups)

$$MS_A=\frac{SS_A}{df_A} \tag{4.19}$$

$$MS_V=\frac{SS_V}{df_V} \tag{4.20}$$

$$MS_{AV}=\frac{SS_{AV}}{df_{AV}} \tag{4.21}$$

Total mean of squares

$$MS_{total}=\frac{SS_{total}}{N-1} \tag{4.22}$$

Mean square (within group)

$$MS_w=\frac{SS_w}{f_w} \tag{4.23}$$

Variance ratio (F-ratio)

The F-ration (statistical) defines as the ratio of variance between groups and variance within groups.

$$F_A=\frac{MS_A}{MS_w} \tag{4.24}$$

$$F_V=\frac{MS_V}{MS_w} \tag{4.25}$$

$$F_{AV}=\frac{MS_{AV}}{MS_w} \tag{4.26}$$

4.2.3 Analysis decision

To take the decision about confirmation of the hypothesis there must be compared statistic variance ratio with F-distribution[17]. F-distribution is also considered as critical F-ratio. For example, if F-statistic more than F-critical then it means that with the set significance the null hypothesis is rejected. The null hypothesis is confirmed in a case if F-statistic is equal to F-distribution.

5 Analysis

5.1 General test feedback

The measurement was successfully accomplished. Thus the test had been launched twice. After the first launch of the measurement the feedback from subjects was rather negative, so it was taken a decision to change the conditions and launch the test again. The main negative factor was the length of the test. In spite of the actual significant length 26 minutes, overall test with explanations of test conditions and tutorial example took almost 30 minutes. That caused subjects' fatigue and decreasing of evaluation adequacy by the end of the measurement. The distinction between two launches was division the second measurement into two parts. After division the feedback of the subjects increased significantly. In spite of overall positive feedback all the subjects mentioned rather often stimuli repetition, but it was one of the principal test conditions. Referring to after-test subjects' response the noticeable preference between all the 5 original audiovisual sequences were given to the video with swans, the least favourite sequence was "cars" video. The subjective preference of the "swans" video was based on overall aesthetic feeling caused by relaxing video and audio content. The objective results are described in the next part of the current chapter

As for researcher the most complicated part in the testing was an organisation of subjects, especially in the second launch of the measurement. Additional difficulties were caused by organisation of participation for the second part of the measurement.

After successful test accomplishment data were carefully transferred into data table. For the data analysis were taken results only from the second launch. The next and conclusive step after those is data analysis.

5.2 Average evaluation results

Based on test results were made two graphs: one is sorted by video quality (Fig 5.1) and the second one by audio quality (Fig 5.2.). Comparing two graphs it is clearly visible that video quality is more essential for overall audiovisual perception.

Form the objective results it is clearly visible, that the "swan" video is preferred till the video quality decreases from 2Mbps to 1Mbps. It can be concluded, that till that limit the video

quality distortion are insignificant. The same perception breakdown is presented on a "wall" video, where video quality is more significant.

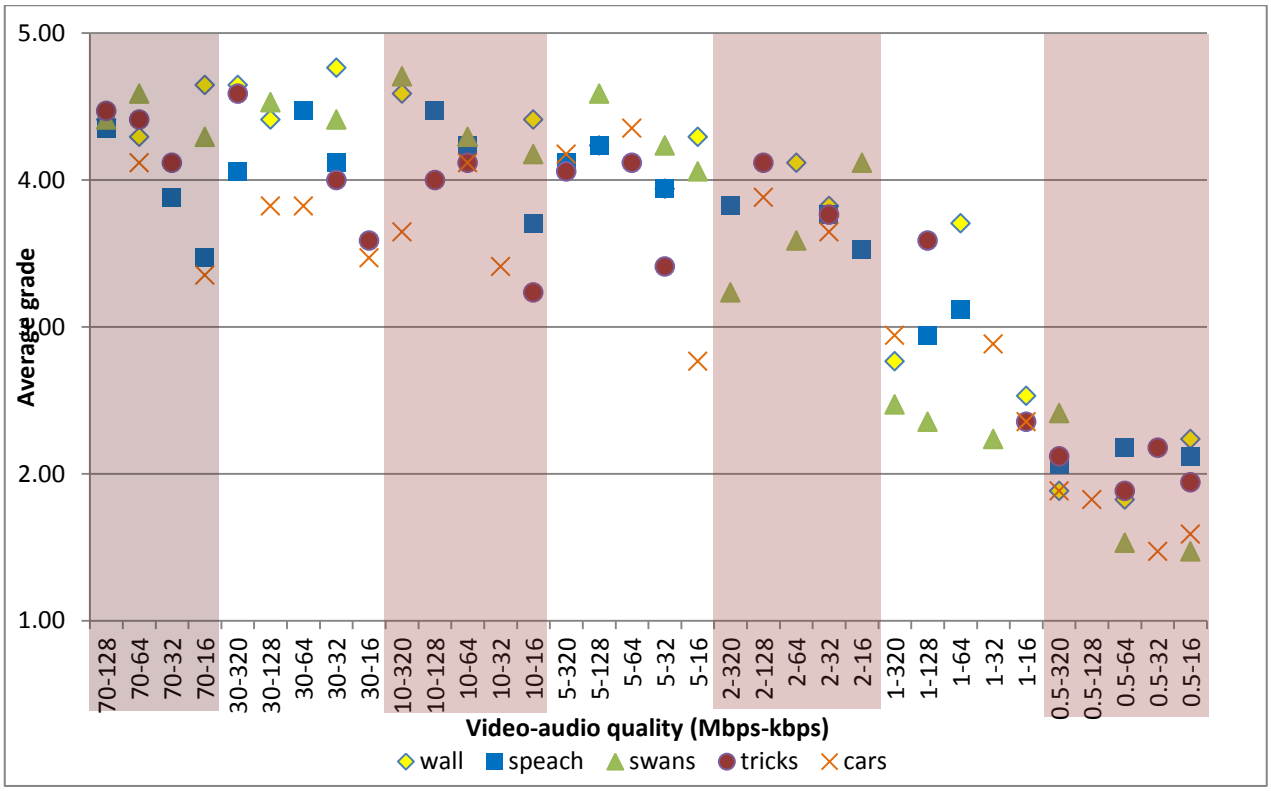


Fig 5.1 Average evaluation results (by video quality)

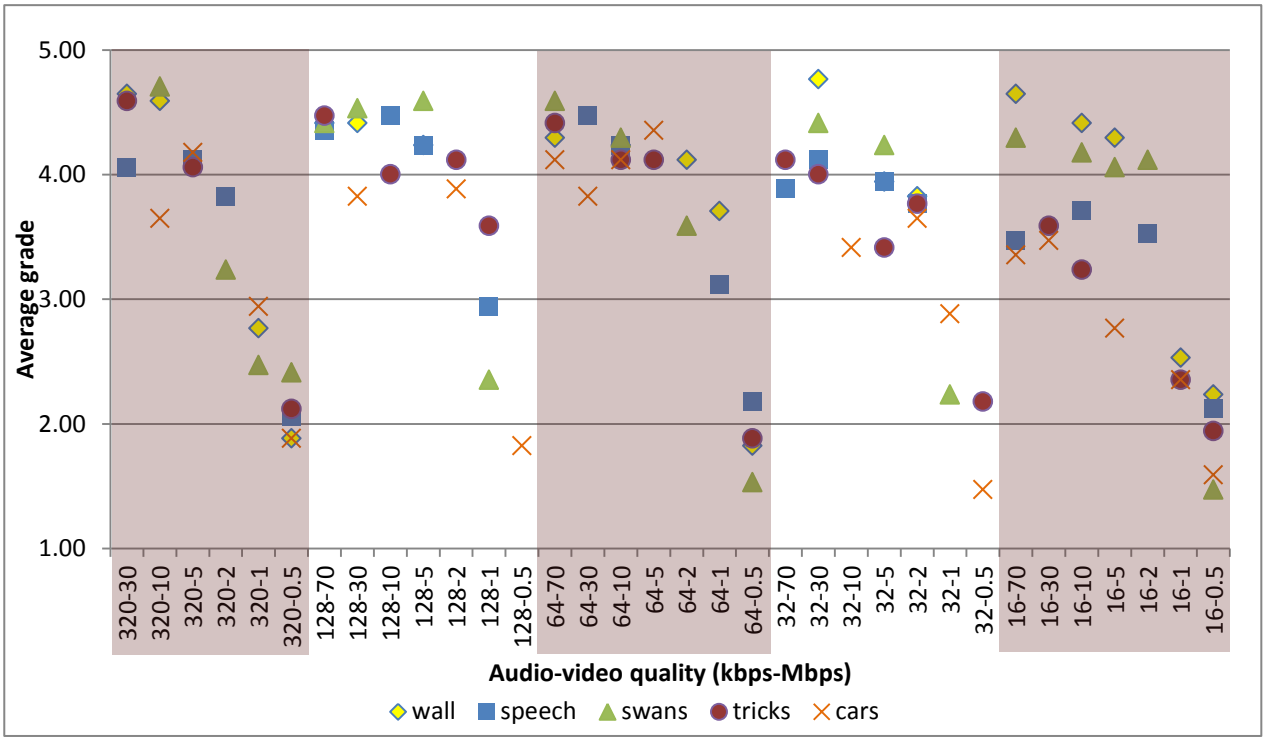


Fig 5.2 Average evaluation results (by audio quality)

Below are depicted 5 graphs (Fig 5.3-5.8) with mean values for each content.

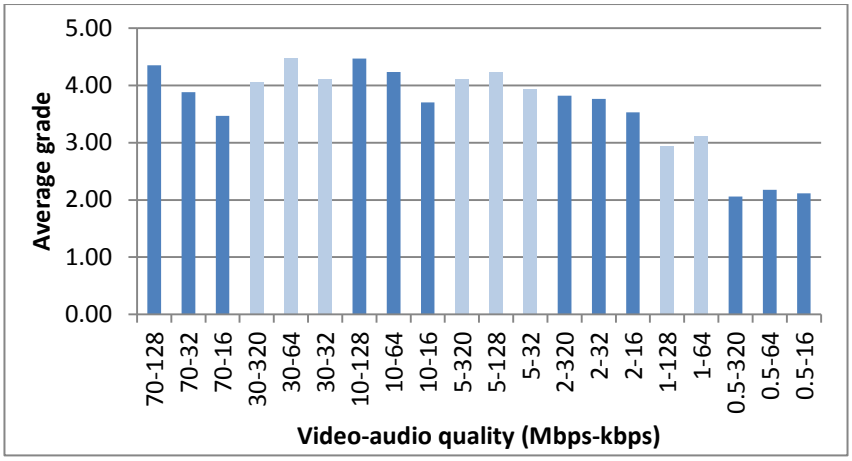


Fig 5.3 Average evaluation results for the "Speech" content

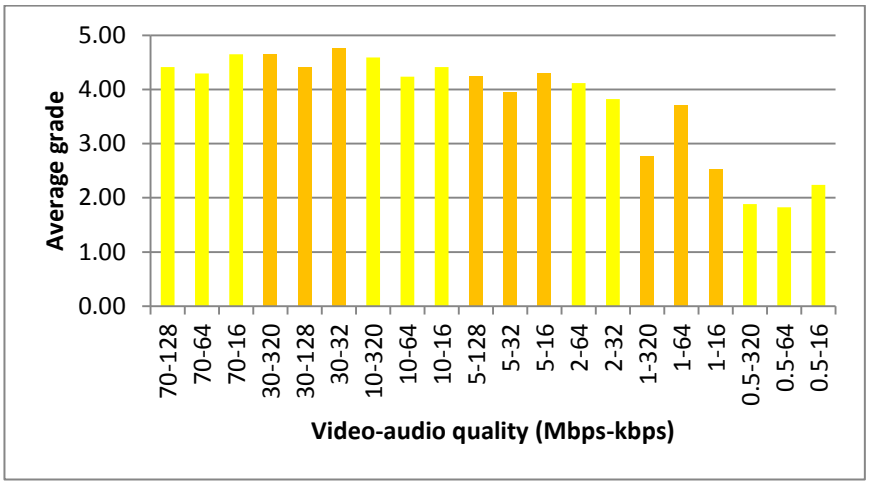


Fig 5.4 Average evaluation results for the "Wall" content

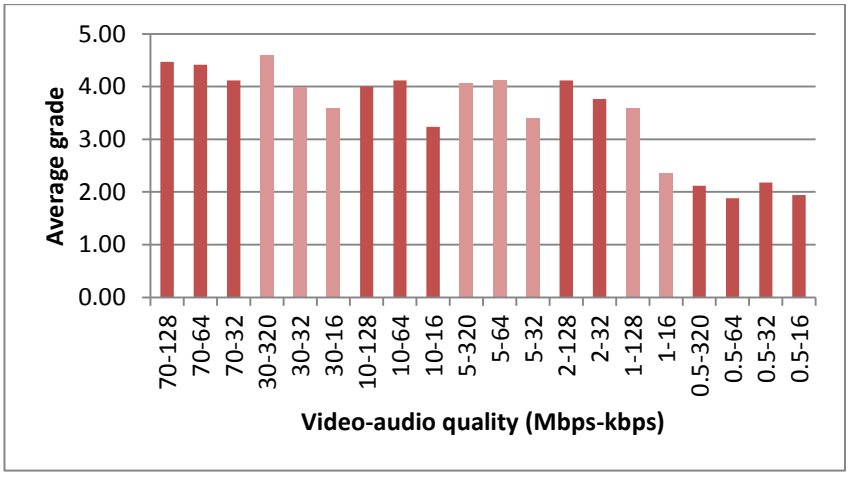


Fig 5.5 Average evaluation results for the "Tricks" content

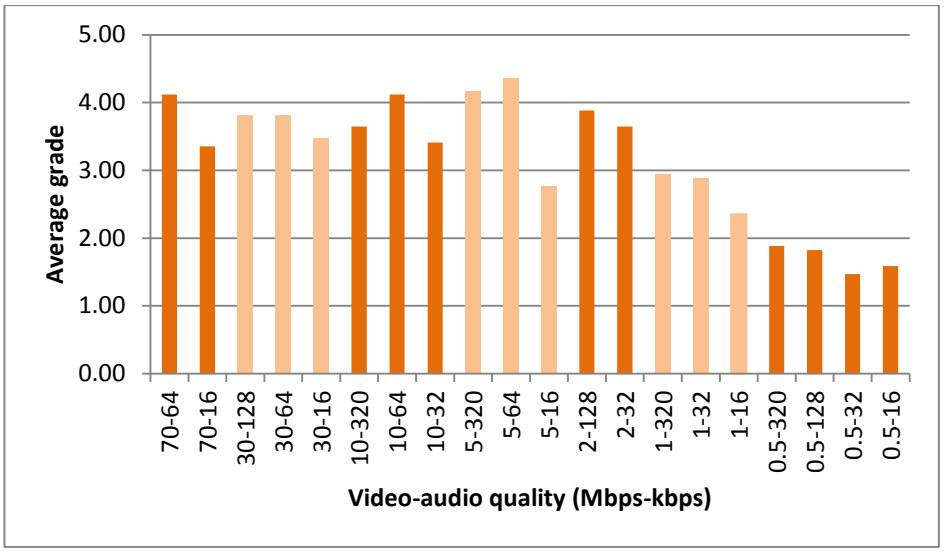


Fig 5.6 Average evaluation results for the "Cars" content

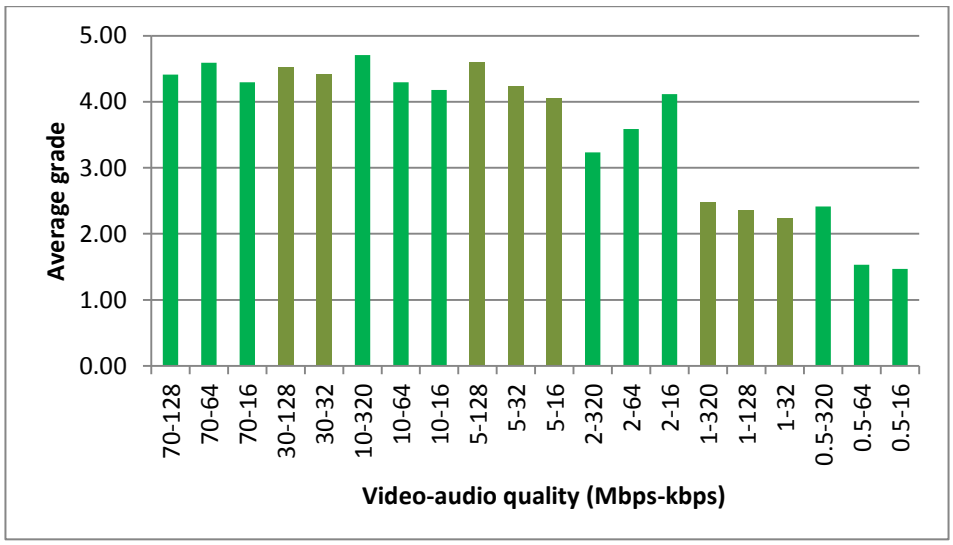


Fig 5.7 Average evaluation results for the "Swans" content

The graphs above represent that for different contents means distributed differently. The more significant is difference in evaluations of audio degradation within the same video quality range the more influence has audio quality to overall perception. The general range of evaluation grades depends on aesthetic parameters of the audivisual sequences.

The more precise analysis will be discussed in following parts of the capter with results of T-test and ANOVA.

5.3 Interpretation of T-test results

The main function of a T-test is to determine the difference between mean values of two test groups. For the particular task the groups were defined as different contents with the same quality parameters. Thus there were formed 10 pairs: speech/wall, speech/tricks, speech/cars, speech/swans, wall/tricks, wall/cars, wall/swans, tricks/cars, tricks/swans, cars/swans. The results of T-test are depicted in Fig 5.8.

The result of the T-Test can be concluded by acceptance or rejection of one of two hypotheses:

H_0: Content of audiovisual sequences has no significant influence on overall perception

H_A Content of audiovisual sequences has significant influence on overall perception

Acceptance or rejection of hypotheses depends on significance level, which is usually is equal to 5%. The hypothesis is considered to be rejected if the probability of means conformity is less than 0.05. At the graph (Fig 5.8) the significance level is indicated with a red line.

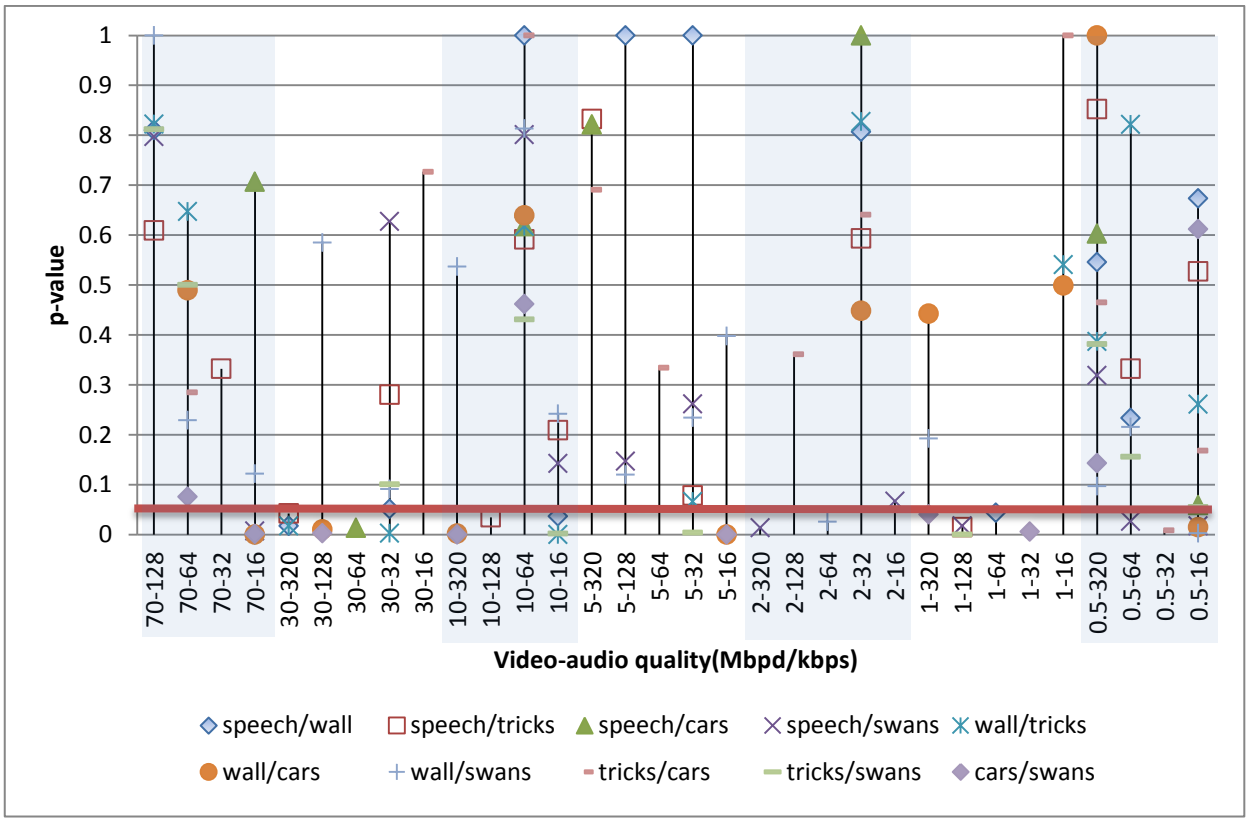


Fig 5.8 T-test results

Since T-test implies analysis of only two parameters, in the following paragraphs it is going to be discussed results of each pair comparison.

Speech/wall

The means analysis showed that zero hypotheses is accepted in 62% of possible combinations (Appendix 5a). Both of the contents has a sufficient audio ("Speech" sequence, due to important audio information) or video ("Wall" sequence, due to particularities in the picture) parameter. An apparent difference appears when these two parameters are set opposite (the highest video quality and the lowest audio quality). Thus the likeliness probability is higher when parameters are set equally (the highest/lowest video quality and the highest/lowest audio quality).

Speech/swans

For a speech/swans combination the hypothesis of equal evaluation for any content is accepted in 54% of cases (Appendix 5b). By degradation of video qualities the evaluations of the "Speech" sequence decreases rather uniformly, whereas the "Swans" sequence has an obvious breakdown at the point of dropping video quality from 2Mbps till 1Mpbs. The "Swan" sequence does not contain any important information in audio, thus the video quality is sufficient. In this case zero hypothesis is rejected when the one of parameters achieved it's extremely value.

Speech/tricks

The likeliness of these two factors is around 78% (Appendix 5c). A fast moving video sequence with tricks is supported with distinct musical audio track with lyrics, thus the audiovisual sequence contains important information in both audio and video. The sequence with lyrics got higher evaluation when the highest audio quality was set. When the audio quality was set to minimal the "Speech" sequence received higher rate than the "tricks" one. It can be concluded that with low audio bitrate video sequence additional sounds to speech generates additional distortion to overall perception.

Speech/cars

In the case of speech/car pair the zero hypotheses is accepted in 71% of cases (Appendix 5d). Despite the "Car" sequence does not contain any significant information in audio; perception of noise distortion as speech distortion is sensitive to audio bitrate degradation.

Wall /swans

The probability of evaluation similarity of “Swans” and “Wall” sequences is 87% (appendix 5e). Both of these sequences contain sufficient information in video part and additional information in audio part (both of them are accompanied with distinct audio tracks without lyrics). Thus the video of “Wall” sequence contains more details. T-test showed that “Swans” and “Wall” sequences can be concerned as the sequences with the same content.

Wall/tricks

The contents of “Wall” and “Tricks” sequences do not influence on overall perception in 66% of cases (appendix 5f). Despite of video quality is sufficient for both sequences; the “Tricks” sequence is more sensitive to audio quality changes. The likeness of means is more probable when video parameter is set to the minimum not depending on audio quality. The difference is more significant when a high audio degradation.

Wall/cars

The zero hypothesis is accepted in 54% of cases (appendix 5g). Due to high dependence of overall perception on the video of the “Wall” sequence, its mean value is greater until the high degradation of video bitrate. The main difference of means was noticed in the quality set where the audio parameter is set to the minimum. Thus it is proved again that distortion of ambience noise in the “Cars” sequence is significant for overall perception.

Tricks/cars

The similarity of “Tricks” and “Cars” sequences is probable in 90% of cases (appendix 5h). That can be explained with the similar perception of audio degradation in sequences containing speech and ambient noise in addition to high motion picture.

Cars/swans

The least probability of means similarity was noticed in the combination of “Cars” and “swans” sequences – only 30% (appendix 5i). The similarity is more probable when the degradation of both audio and video parameters is high. In these cases the probability of content dependence on overall perception is highly noticed.

Tricks/swans

The equal content perception is presented in 60% of cases with “Tricks” and “Swans” audiovisual sequences (appendix 5j). The more audio quality is degraded the more different are means of the sequences.

5.4 T-test conclusion

The results of T-test showed that overall perception of an audiovisual sequence is dependent on a content of this sequence. Influence of audio or video quality on overall perception is also dependent on the content, what makes it more sensitive to either video either audio changes.

Concerning to audio content it was noticed, that speech, background music with lyrics, and ambient noise perceived similarly, even if only the speech contains significant information. With the same rather high video quality the overall perception of audio-dependent sequences visibly degrades with decreasing of audio quality, the lower is a video bitrate the less audio degradation influences on overall perception.

If video-dependent contents contain more sufficient information, when decreasing video quality the overall perception decreasing uniformly until a breakdown point, after particular degradation the perception drops impetuously.

5.5 Interpretation of ANOVA results

ANOVA analysis also implies an acceptance or rejection of set hypotheses. For the presented analysis were set the following hypotheses:

H_0: Overall perception of audiovisual sequences doesn’t depend on the content of the sequences.

H_A: Overall perception of audiovisual sequences depends on the content of the sequences.

Each set of hypotheses was applied for every set of audio-video qualities; for the test were used 32 quality sets. The conclusion of hypothesis acceptance/rejection is dependent on F-distribution, the larger is difference between F-ratio and F-critical, the more is difference between variances. ANOVA unlike the T-test can analyse more than 2 groups of variances. Below in Fig 5.9 is depicted the result of ANOVA for 5 groups (dependence between a quality set and the contents of audiovisual sequences).

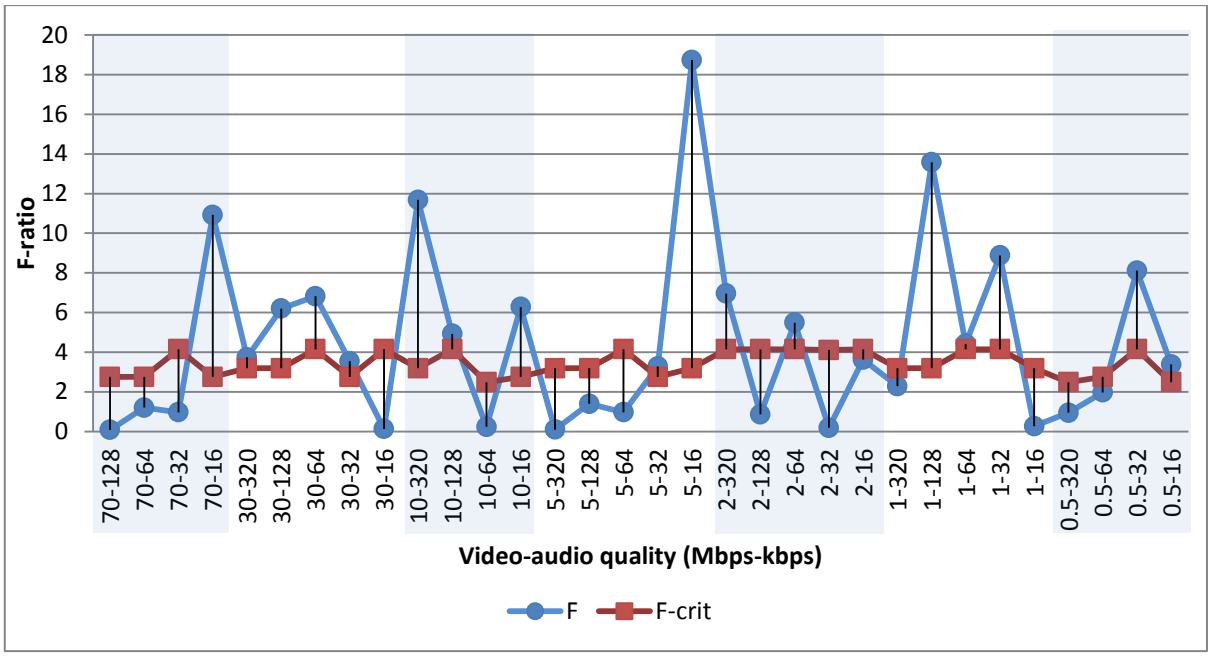


Fig 5.9 Dependence between F and F-critical ratios

The graph is clearly represents, that the audiovisual perception in most of audiovisual parameters sets is dependent on the content of the sequences. Thus the higher difference between variances presents when audio degradation is sufficient.

5.6 Analysis of contents

Analysis of the influencing parameters on overall quality of audiovisual sequences is summarised in a table 12. The table 12 is an accomplished variation of table 6 and summation of T-test and ANOVA results.

	Short description		Influence on overall perception	
	Video	Audio	Video degradation	Audio degradation
"Wall"	High detailed slow-motioned	Background music without lyrics	Sufficient, breakdown point	Not sufficient
"Speech"	Insignificant motion	Speech containing significant information	Not sufficient	Sufficient
"Swans"	Slow motioned big objects, stable background	Background music without lyrics	Breakdown point is sufficient	Not sufficient
"Tricks"	High motion object, almost stable background	Background music with lyrics	Breakdown point is sufficient	Sufficient
"Cars"	High motioned big oncoming objects	Ambient noise	Breakdown point is sufficient	Sufficient

Table 12 Influence of audio and video degradation on overall perception

CONCLUSION

The aim of the presented research was to find an interaction between audiovisual quality and the content of the sequences by a statistical analysis of subjective test results.

Before started the actual measurement there were done a preparatory research on previous works with resembling themes, to conclude the achieved results and use them as background information for the thesis. Also there were studied different possible psychometric methods of testing and statistical analysis methods. That was the theoretical part of the thesis.

For the practical part of the research there were created 5 original audiovisual sequences which were used for further compression and audiovisual quality degradation. Thus were created 135 audiovisual sequences with variable audio and video quality. The audio qualities had 5 levels of degradation and the video qualities had 7 levels of degradation. As test stimuli were taken 100 random audiovisual sequences (20 from each content).

The stimuli base prepared for the measurement implied a large number of evaluations. That is why the test was designed using rather simple psychometric method of successive categories. The corresponding evaluation method allowed to receive appropriate results using the limited time and not reducing the number of stimuli.

The test took place in an audio measurement booth, to provide all the subjects with equal audiovisual conditions. The measurement was divided into two parts by 15 minutes. The overall feedback of participated subjects about the test was rather positive; the only disadvantage of the test was the limited number of the original contents. However the number of contents was one of the applied test methods, and could not be increased.

After all the subjects successfully finished the test, data were entered to the database and analysed using T-test and ANOVA. The general analysis of mean values showed that for any content audiovisual sequences the sufficient parameter for overall perception is a video quality. Also was noticed a perception breakdown on the video degradation point of 2Mbps, before this point the overall perception has rather similar evaluation level, but after that point the evaluation results significantly decreased.

The more particular analysis of variances showed that for audio-dependent contents the audio degradation is sufficiently decreasing the overall perception when the video quality is rather

high, but after the breakdown point, audio degradation does not influence on the overall perception. That can be concluded that the increasing of audio bitrate for audiovisual sequences with low video bitrate has no influence on overall perception.

Also during the work was done one interesting conclusion, that the degradation of speech, songs with lyrics and ambient noise are perceived equally.

The further development of this theme can contain more particular research on the studying the breakdown point of different contents and as the second division of the theme to study the physical reasons of equal perception of voice and ambient noise audio sequences.

References

[1] GOFF, D.: Fiber Optic Video Transmission, Focal Press, 2013,

[2] HANDS D.S.: A Basic Multimedia Quality Model, IEEE transactions on multimedia, vol. 6, no. 6, 2004.

[3] LASSALLE,J.: Impact of the content on subjective evaluation of audiovisual quality" IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 27-29 june 2012, Seoul, Rwpublic of Korea, 2012.

[4] RIES, M.; GARDLO B.: Audiovisual Quality Estimation for Mobile Video Services, IEEE journal on selected areas in communications, vol.28, no.3, April 2010.

[5] WINKLER,S.; FALLER, C.: Precieved Audiovisual Quality of Low Bitrate Multimedia Content, IEEE transactions on multimedia vol.8, 2006.

[6] WOSZCZYK, W; BECH, S.; HANSEN, V.: "Interactions between audio-visual factors in a home theatre system: definition of subjective attributes", McGiU University, Montreal, Quebec, Canada, 1995.

[7] REITER,U.: Overall perceiver audiovicual quality – what people pay attention to, Norwegian University of Science and Technology, 2011.

[8] GULIFORD, J.,P.:Psychometric Methods, McGRAW-HILL BOOK COMPANY,INC., Second edition, 1954.

[9] ITU-R Recommendation BS. 1534: Method for the subjective assessment of intermediate quality levels of coding systems, 2003.

[10] ITU-R Recommendation BS.1284-1: General methods for the subjective assessment of sound quality, 2003.

[11] ITU-R Recommendation P.911: Subjective audiovisual quality assessment methods for multimedia application,1998

[12] TOMINAGA,T., HAYASHI,T., OKAMOTO J., TAKAHASHI A.: Performance Comparisons Of Subjective Quality Assessment Methods For Mobile Video, NTT Service Integration Laboratories, NTT Corporation, Midori–cho, Musashino–shi, Tokyo, Japan, pp180–8585.

[13] DINHAM, S.M.: Exploring Statistics, Brooks/Cole Publishing Co., pp294

[14] HERRERA, M.: Evaluation of audio coding artefacts, Doctoral Thesis, CVUT, 2009.

[15] WILLIAMS,R.: Two-Way Analysis of Variance [online], University of Notre Dame, Sociology Graduate Statistics I. [cit. 2012-08-13]. Available from WWW: <http://www3.nd.edu/~rwilliam/stats1/x61.pdf>.

[16] MICHALIK,P.: Protokol O Měření Po Realizaci Měřící Komory (CVUT,FEL,Dejvicka), SONING, 17.4.2010

[17] GRAZIANO,A., RAULIN, M.: Research Methods, State University of New York at Buffalo 6th edition, 2006

APPENDIXES

Appendix 1 **Experimental audio-visual sequences set**

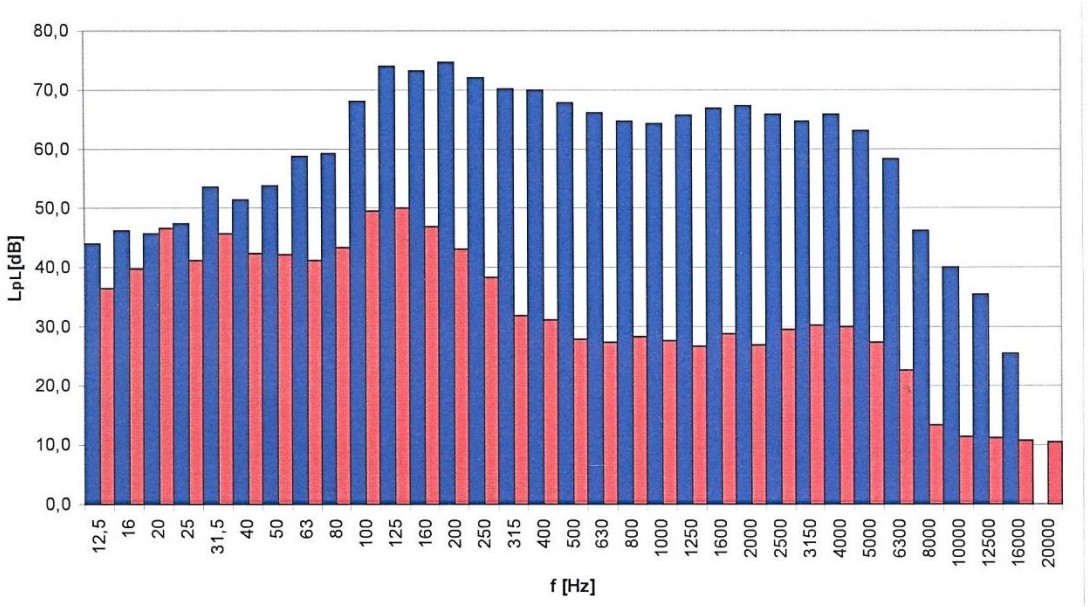
	set 1	
№	video name	duration, sec
1	4-70-32	12
2	2-0.5-16	13
3	5-1-320	9
4	3-0.5-320	10
5	1-1-64	10
6	2-30-32	13
7	5-10-64	9
8	4-30-320	12
9	1-0.5-16	10
10	3-2-16	10
11	2-70-32	13
12	4-10-128	12
13	5-30-64	9
14	3-5-128	10
15	1-30-32	10
16	5-10-320	9
17	4-2-128	12
18	3-10-64	10
19	2-1-64	13
20	4-5-32	12
21	5-5-16	9
22	1-5-128	10
23	3-1-32	10
24	4-1-16	12
25	5-0.5-128	9
26	2-0.5-320	13
27	1-5-32	10
28	4-0.5-320	12
29	3-30-128	10
30	2-0.5-64	13
31	5-10-32	9
32	1-0.5-320	10
33	3-70-16	10
34	4-30-32	12
35	5-30-128	9
36	2-10-16	13
37	3-0.5-64	10
38	1-70-128	10
39	4-0.5-16	12
40	5-0.5-32	9
41	3-70-128	10
42	2-5-128	13
43	5-0.5-16	9
44	1-70-64	10
45	4-1-128	12
46	2-2-16	13
47	5-5-320	9
48	4-0.5-32	12
49	2-30-320	13
50	1-10-64	10

	set 1	
№	video name	duration, sec
51	4-5-320	12
52	3-30-32	10
53	5-1-16	9
54	2-1-128	13
55	1-5-16	10
56	4-5-64	12
57	5-0.5-320	9
58	3-5-16	10
59	1-0.5-64	10
60	2-2-32	13
61	4-10-64	12
62	5-1-32	9
63	3-2-320	10
64	4-10-16	12
65	1-2-64	10
66	2-10-128	13
67	5-2-128	9
68	3-0.5-16	10
69	2-30-64	13
70	4-30-16	12
71	1-1-320	10
72	2-70-16	13
73	4-70-64	12
74	1-10-320	10
75	5-5-64	9
76	3-1-128	10
77	1-70-16	10
78	2-70-128	13
79	3-1-320	10
80	4-0.5-64	12
81	5-70-16	9
82	1-30-128	10
83	3-2-64	10
84	5-30-16	9
85	2-5-32	13
86	4-70-128	12
87	3-10-320	10
88	1-1-16	10
89	2-10-64	13
90	5-2-32	9
91	3-5-32	10
92	1-30-320	10
93	2-5-320	13
94	4-2-32	12
95	3-70-64	10
96	1-10-16	10
97	2-2-320	13
98	5-70-64	9
99	3-10-16	10
100	1-2-32	10

Where: *N*-*-* - video number
*-*N*-* - video bitrate, Mbps
--*N* - audio bitrate, kbps

№	**name**
1	Wall
2	Speech
3	Swan
4	Tricks
5	Cars

Appendix 2 **The average energy spectrum of sound pressure level inside and outside the booth**



- out of the booth

- inside the booth

Appendix 3 **Test instruction**

Good afternoon dear observer, thank you for coming!

Today you are going to participate in audio-visual experiment. You will be presented with short audio-visual sequences which you need to evaluate with a scale depicted below:

Grade	**Equivalent**
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

You should take attention to overall audio-visual quality. After each sequence there is a pause for filling up the evaluation table. Please watch it carefully and give an objective opinion.

Appendix 4 **Answer sheet**

Age_____

Video sequence number	Your grade	Video sequence number	Your grade	Video sequence number	Your grade	Video sequence number	Your grade
1		**26**		**51**		**76**	
2		**27**		**52**		**77**	
3		**28**		**53**		**78**	
4		**29**		**54**		**79**	
5		**30**		**55**		**80**	
6		**31**		**56**		**81**	
7		**32**		**57**		**82**	
8		**33**		**58**		**83**	
9		**34**		**59**		**84**	
10		**35**		**60**		**85**	
11		**36**		**61**		**86**	
12		**37**		**62**		**87**	
13		**38**		**63**		**88**	
14		**39**		**64**		**89**	
15		**40**		**65**		**90**	
16		**41**		**66**		**91**	
17		**42**		**67**		**92**	
18		**43**		**68**		**93**	
19		**44**		**69**		**94**	
20		**45**		**70**		**95**	
21		**46**		**71**		**96**	
22		**47**		**72**		**97**	
23		**48**		**73**		**98**	
24		**49**		**74**		**99**	
25		**50**		**75**		**100**	

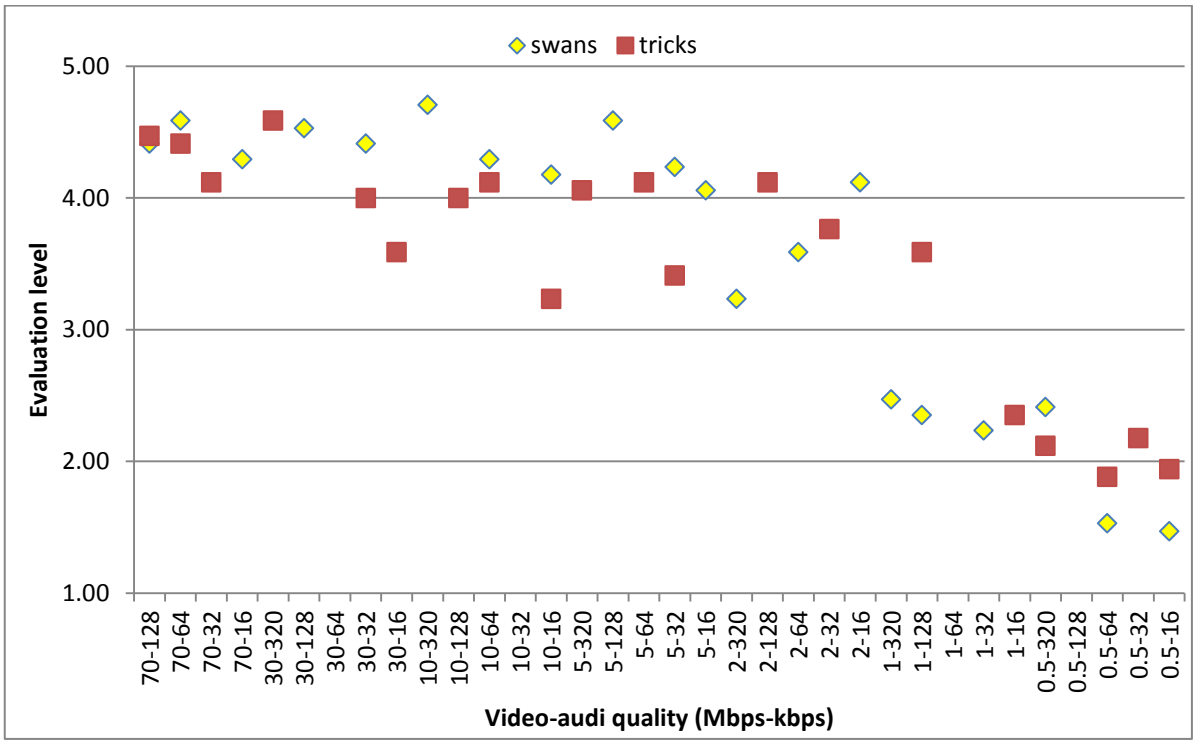
Appendix 5 **Comparison graphs of mean values of two contents sequences and the following T-test result**

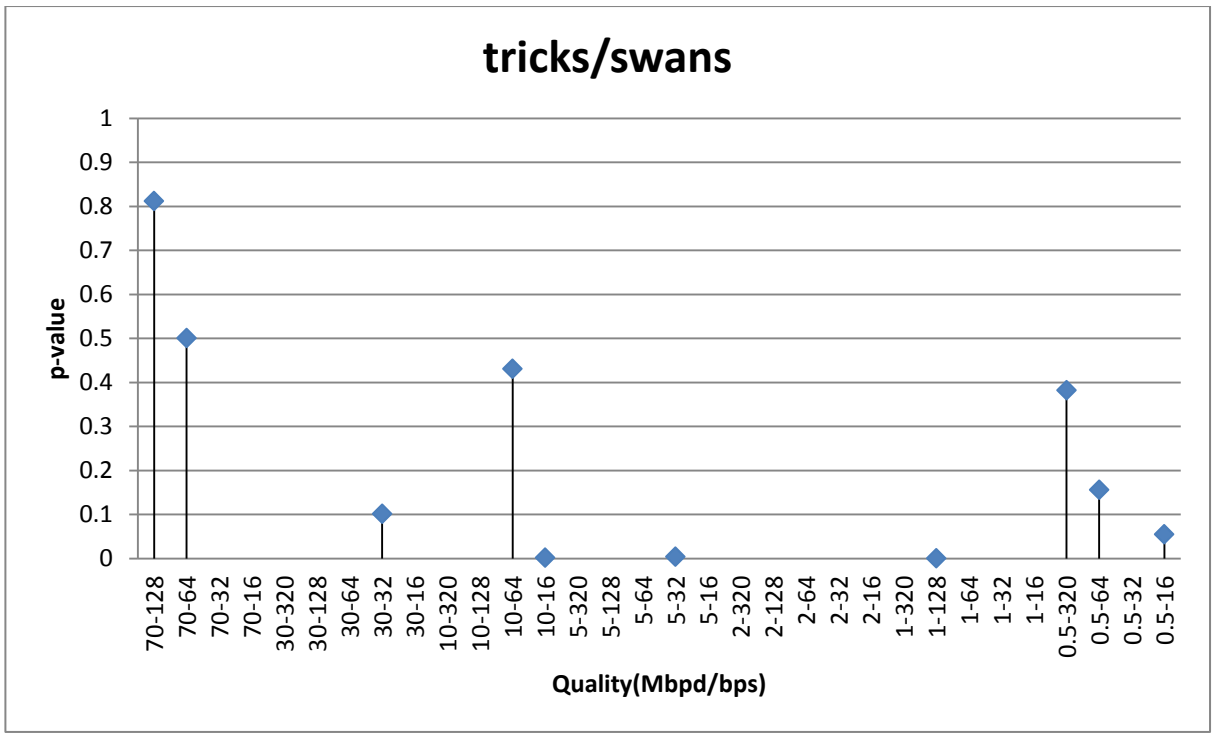
a) Tricks/swans
b) Cars/swans
c) Tricks/cars
d) Wall/cars
e) Wall/tricks
f) Wall /swans
g) Speech/cars
h) Speech/tricks
i) Speech/swans
j) Speech/wall

The first graph represents the mean values of evaluation of two audiovisual signals with different contents. From this graph you can see the influence of audio and video degradation on overall quality.

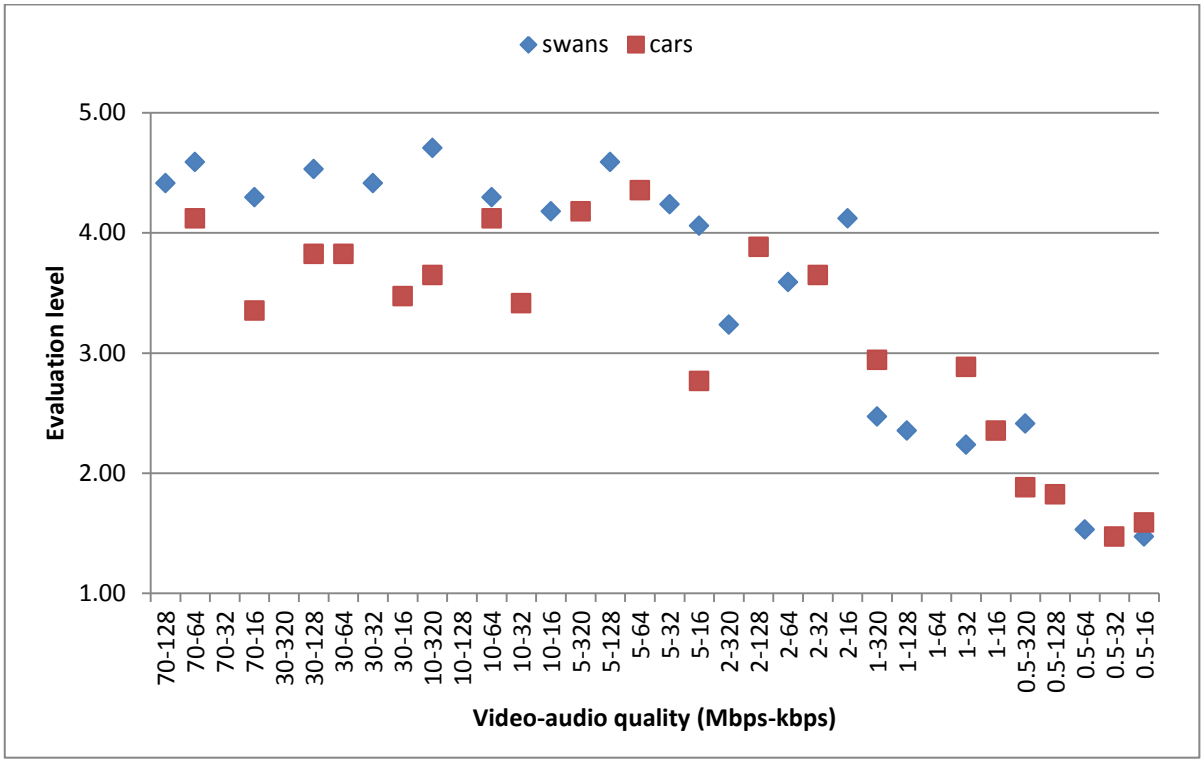
The second graph represents the probability that the mean values of the first signal are equal to a coincident mean value of the second signal.

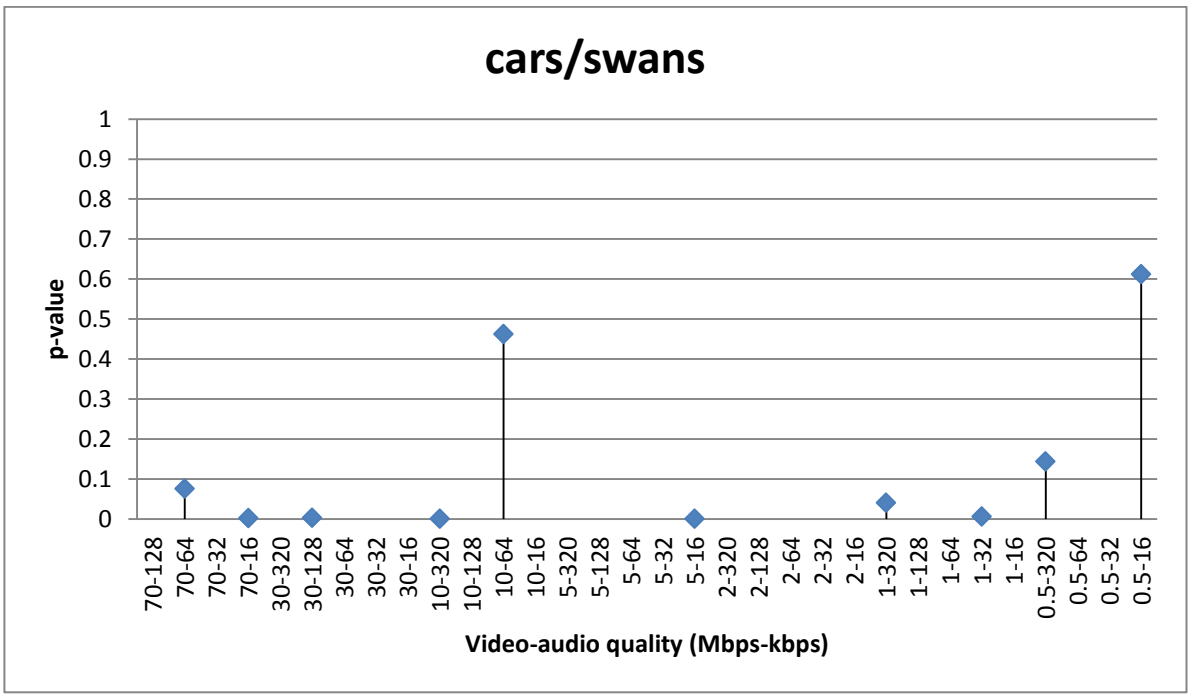
A) Tricks/swans



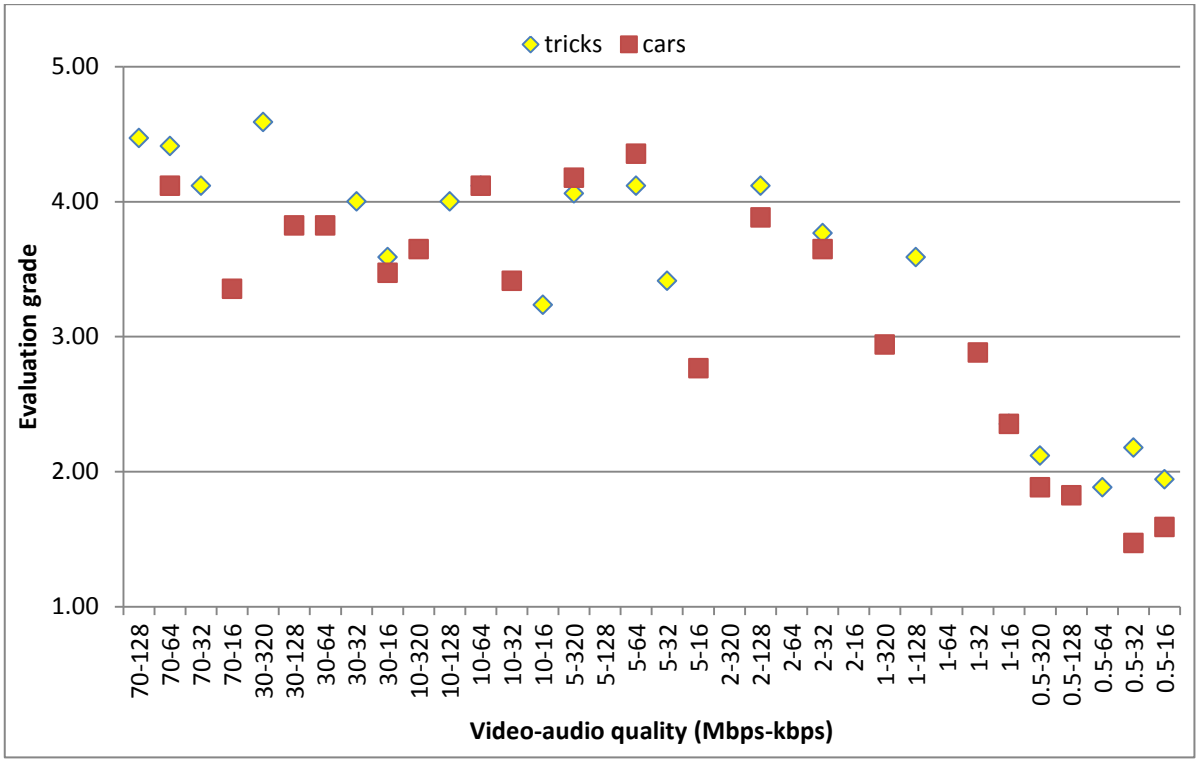


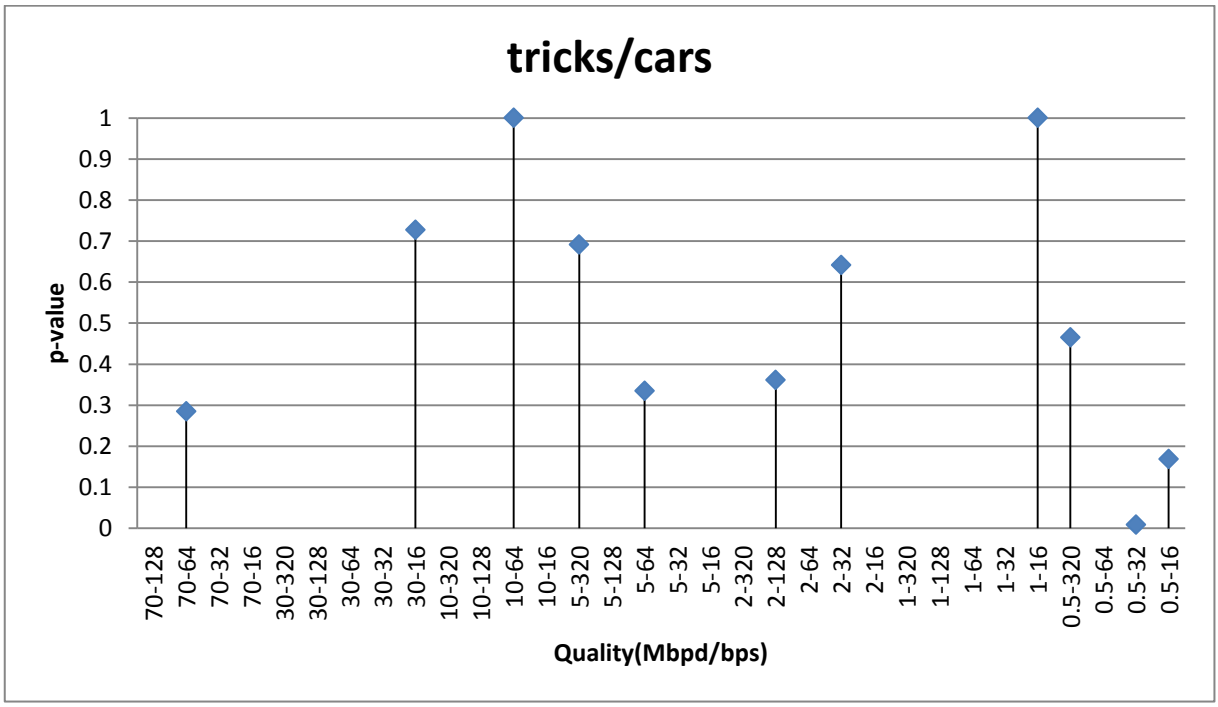
B) Cars/swans


swans
cars
5.00
4.00
3.00
2.00
1.00
Evaluation level
70-128
70-64
70-32
70-16
30-320
30-128
30-64
30-32
30-16
10-320
10-128
10-64
10-32
10-16
5-320
5-128
5-64
5-32
5-16
2-320
2-128
2-64
2-32
2-16
1-320
1-128
1-64
1-32
1-16
0.5-320
0.5-128
0.5-64
0.5-32
0.5-16
Video-audio quality (Mbps-kbps)

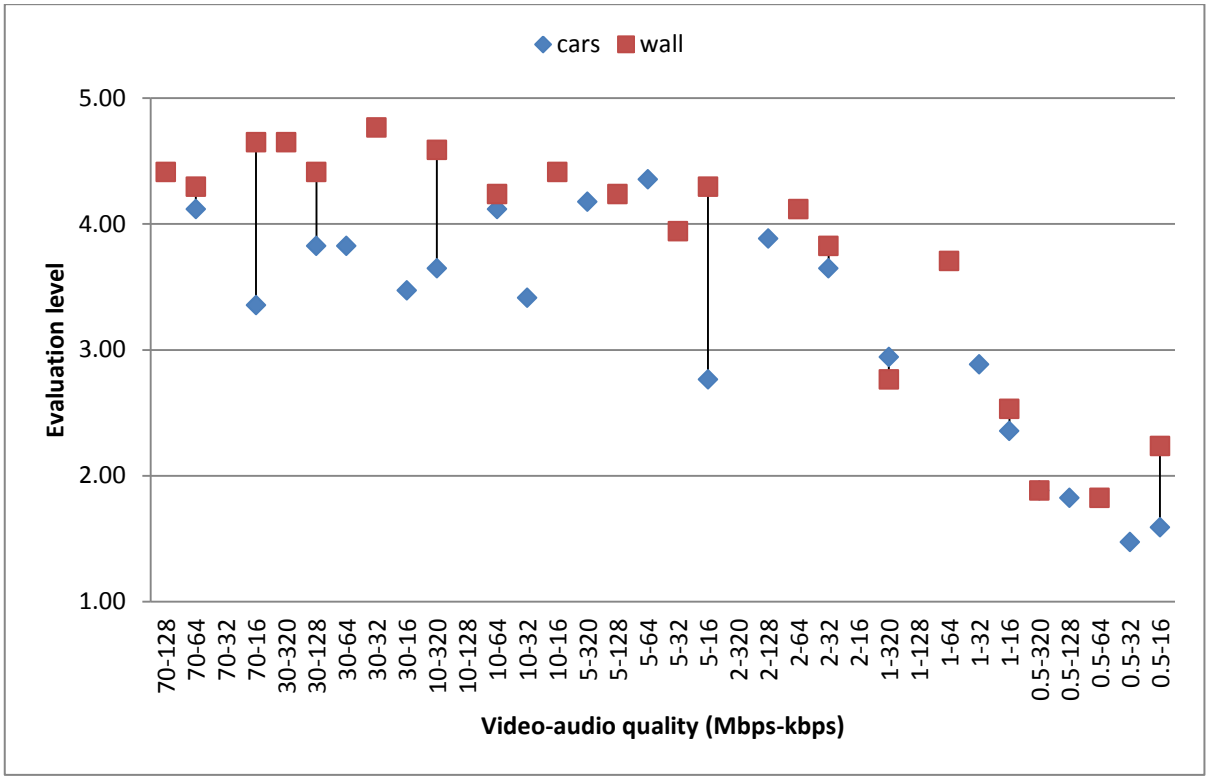

cars/swans
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p-value
70-128
70-64
70-32
70-16
30-320
30-128
30-64
30-32
30-16
10-320
10-128
10-64
10-16
5-320
5-128
5-64
5-32
5-16
2-320
2-128
2-64
2-32
2-16
1-320
1-128
1-64
1-32
1-16
0.5-320
0.5-64
0.5-32
0.5-16
Video-audio quality (Mbps-kbps)

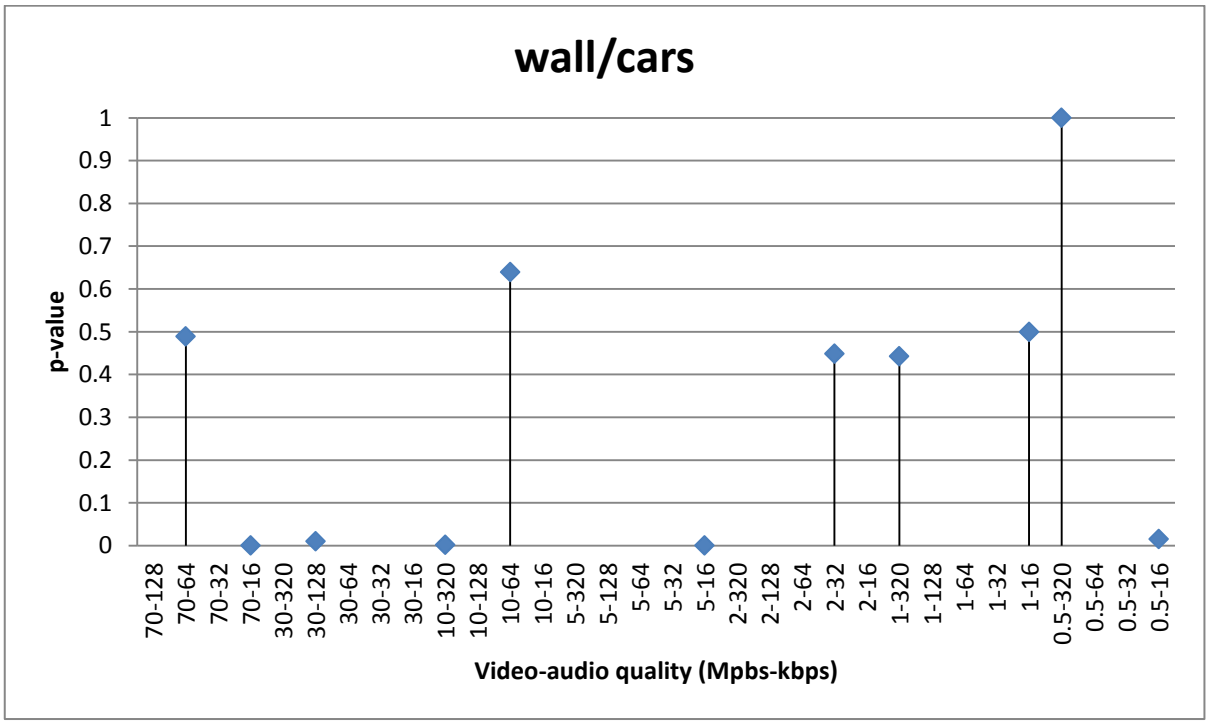
C) Tricks/cars



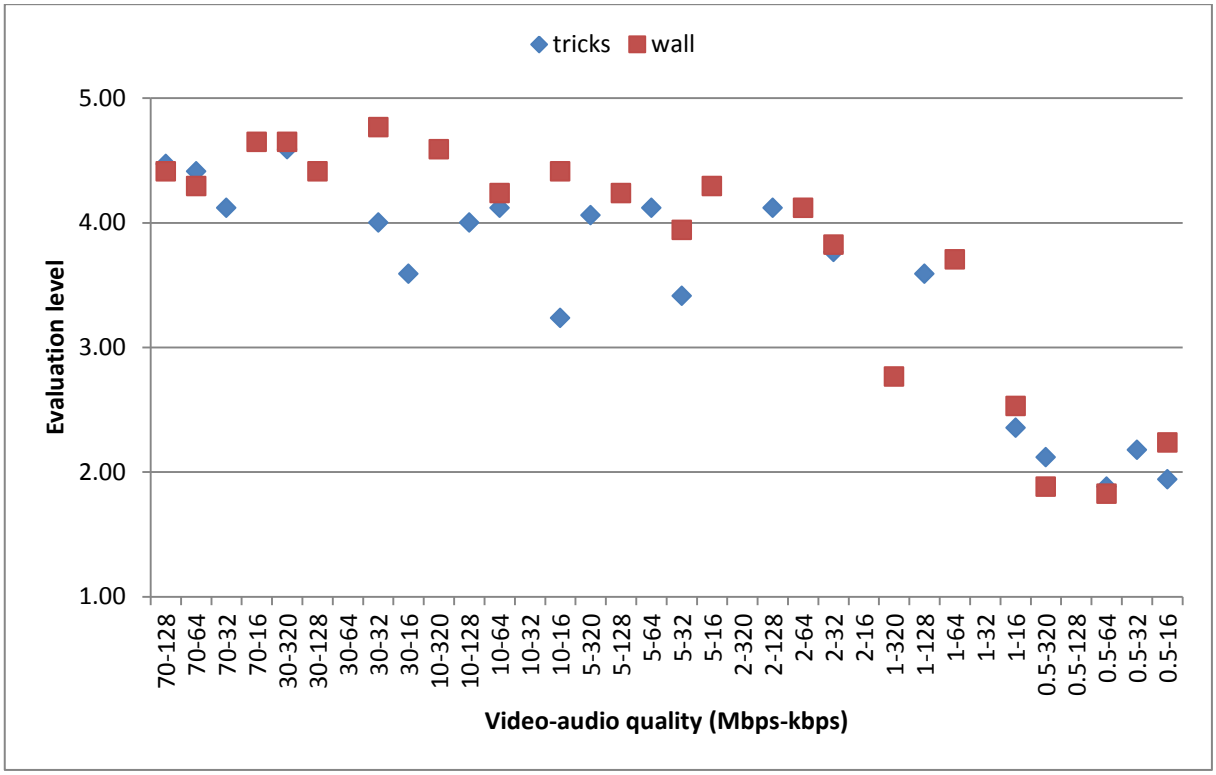


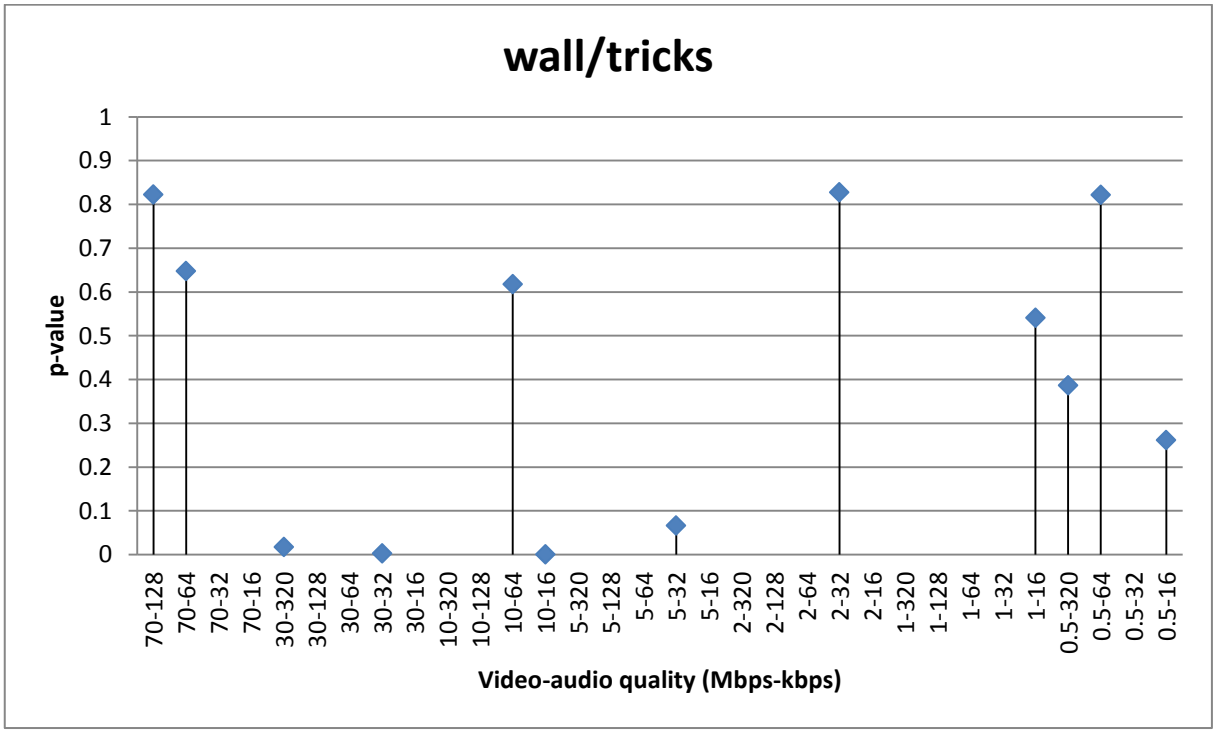
D) Wall/cars


cars
wall
5.00
4.00
3.00
2.00
1.00
Evaluation level
70-128
70-64
70-32
70-16
30-320
30-128
30-64
30-32
30-16
10-320
10-128
10-64
10-32
10-16
5-320
5-128
5-64
5-32
5-16
2-320
2-128
2-64
2-32
2-16
1-320
1-128
1-64
1-32
1-16
0.5-320
0.5-128
0.5-64
0.5-32
0.5-16
Video-audio quality (Mbps-kbps)

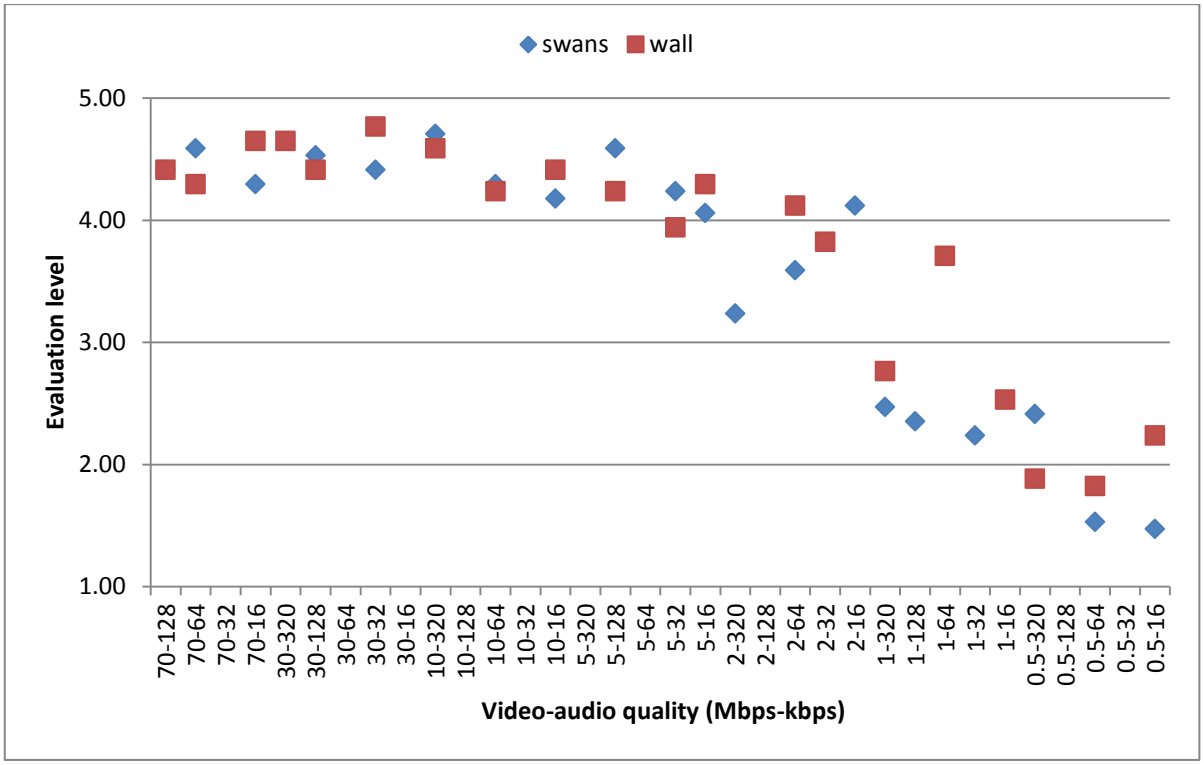

wall/cars
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p-value
70-128
70-64
70-32
70-16
30-320
30-128
30-64
30-32
30-16
10-320
10-128
10-64
10-16
5-320
5-128
5-64
5-32
5-16
2-320
2-128
2-64
2-32
2-16
1-320
1-128
1-64
1-32
1-16
0.5-320
0.5-64
0.5-32
0.5-16
Video-audio quality (Mpbs-kbps)

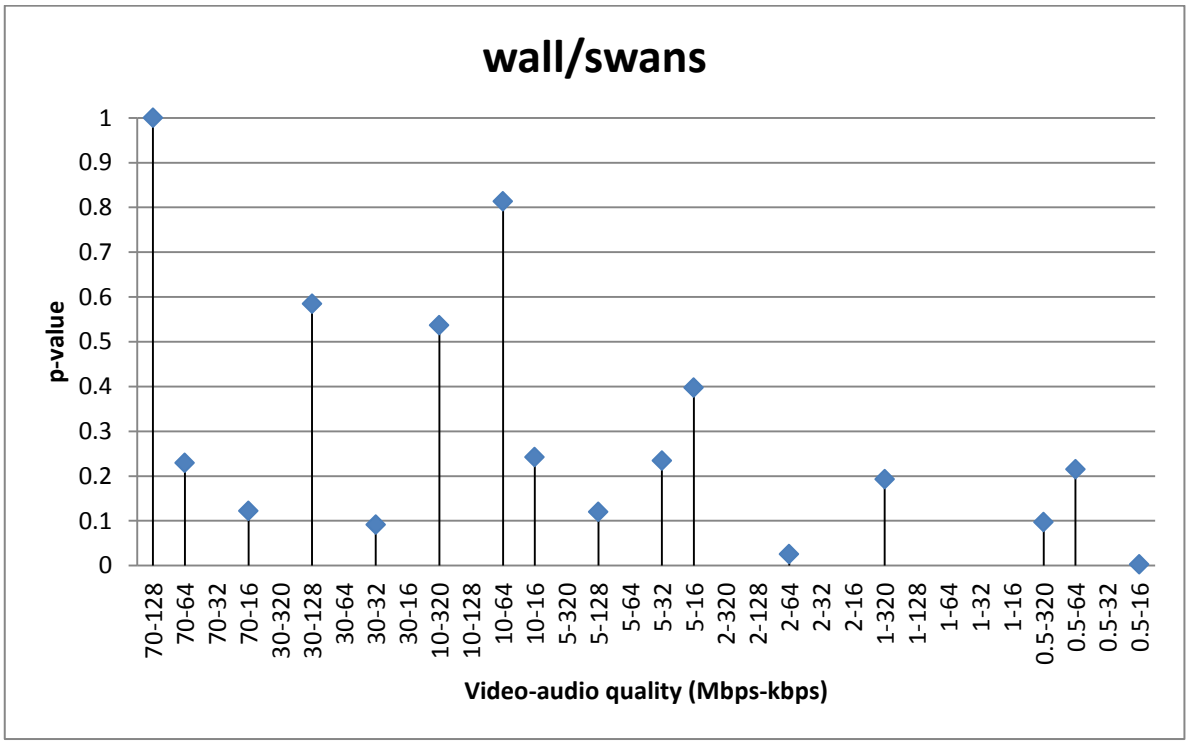
E) Wall/tricks



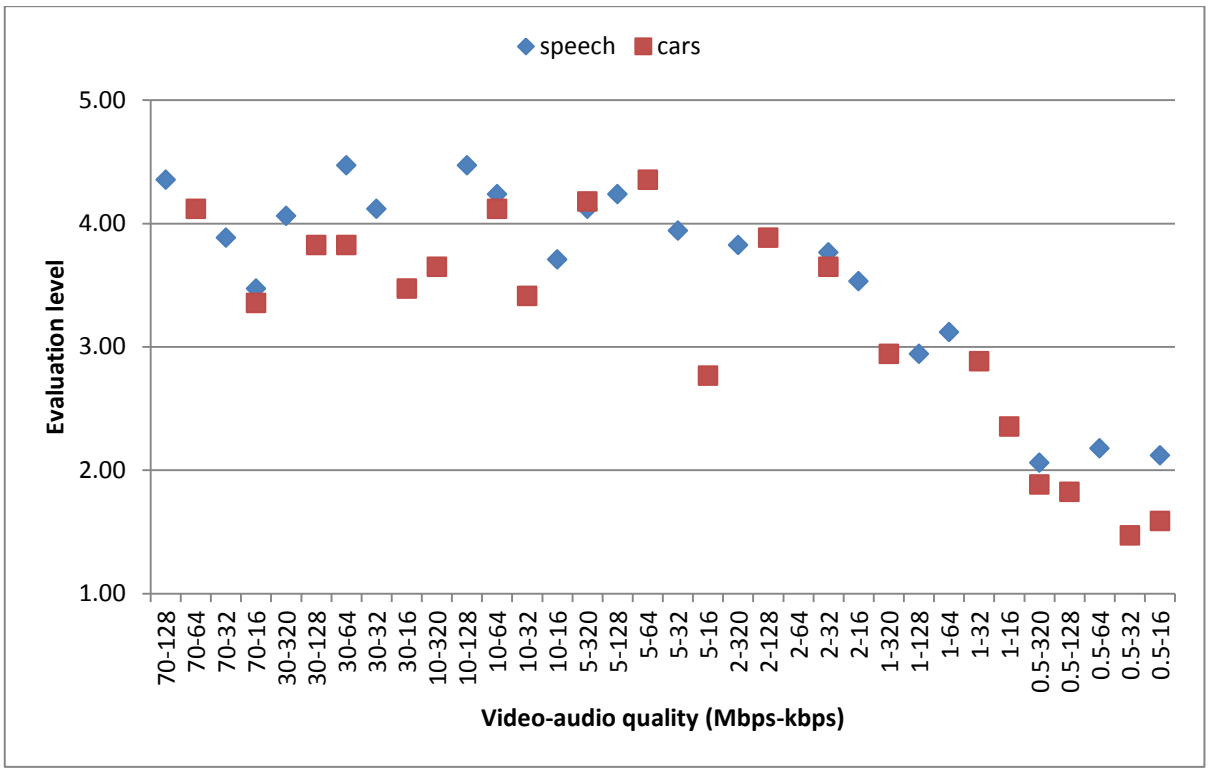


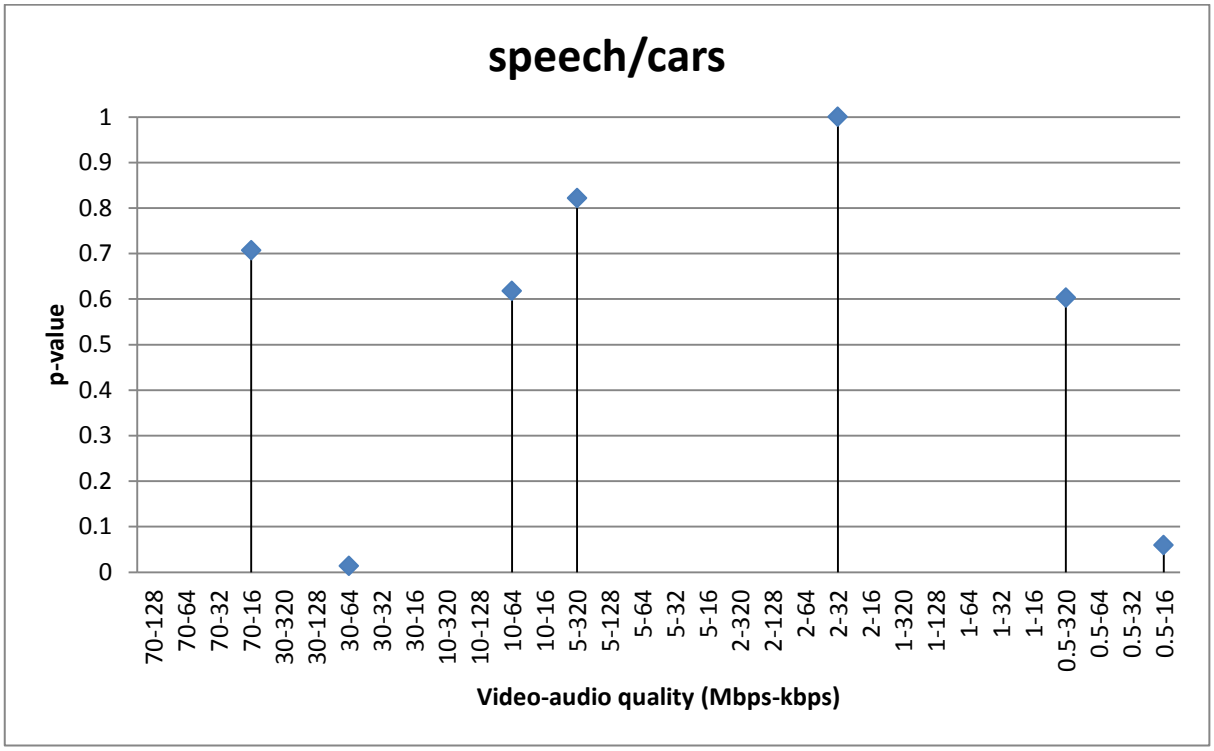
F) Wall/swans


swans
wall
5.00
4.00
3.00
2.00
1.00
Evaluation level
70-128
70-64
70-32
70-16
30-320
30-128
30-64
30-32
30-16
10-320
10-128
10-64
10-32
10-16
5-320
5-128
5-64
5-32
5-16
2-320
2-128
2-64
2-32
2-16
1-320
1-128
1-64
1-32
1-16
0.5-320
0.5-128
0.5-64
0.5-32
0.5-16
Video-audio quality (Mbps-kbps)

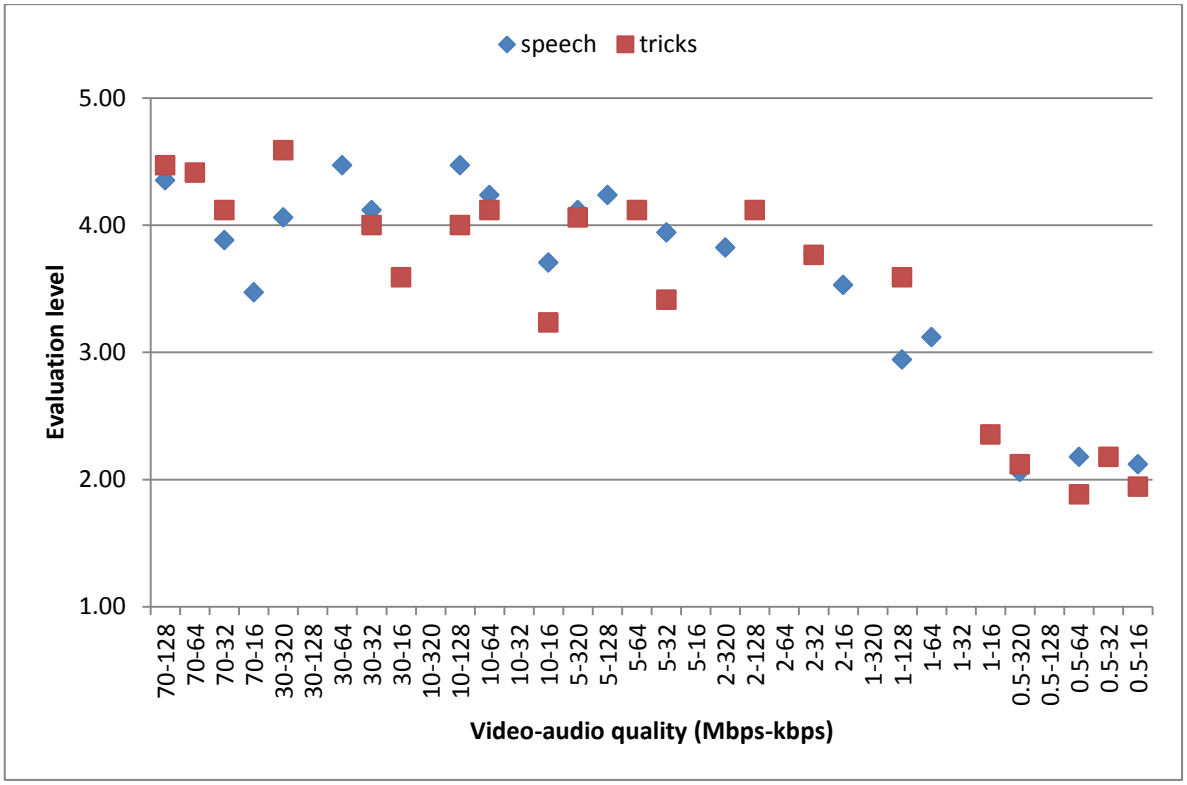

wall/swans
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p-value
70-128
70-64
70-32
70-16
30-320
30-128
30-64
30-32
30-16
10-320
10-128
10-64
10-16
5-320
5-128
5-64
5-32
5-16
2-320
2-128
2-64
2-32
2-16
1-320
1-128
1-64
1-32
1-16
0.5-320
0.5-64
0.5-32
0.5-16
Video-audio quality (Mbps-kbps)

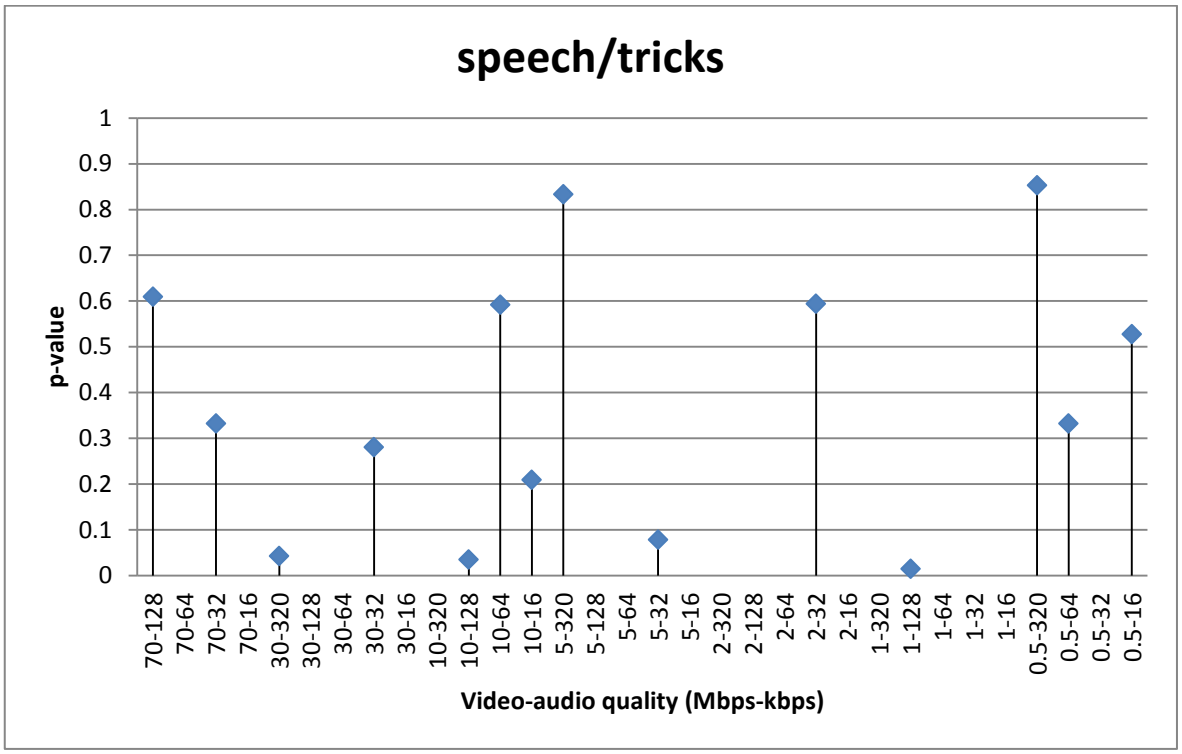
G) Speech/cars



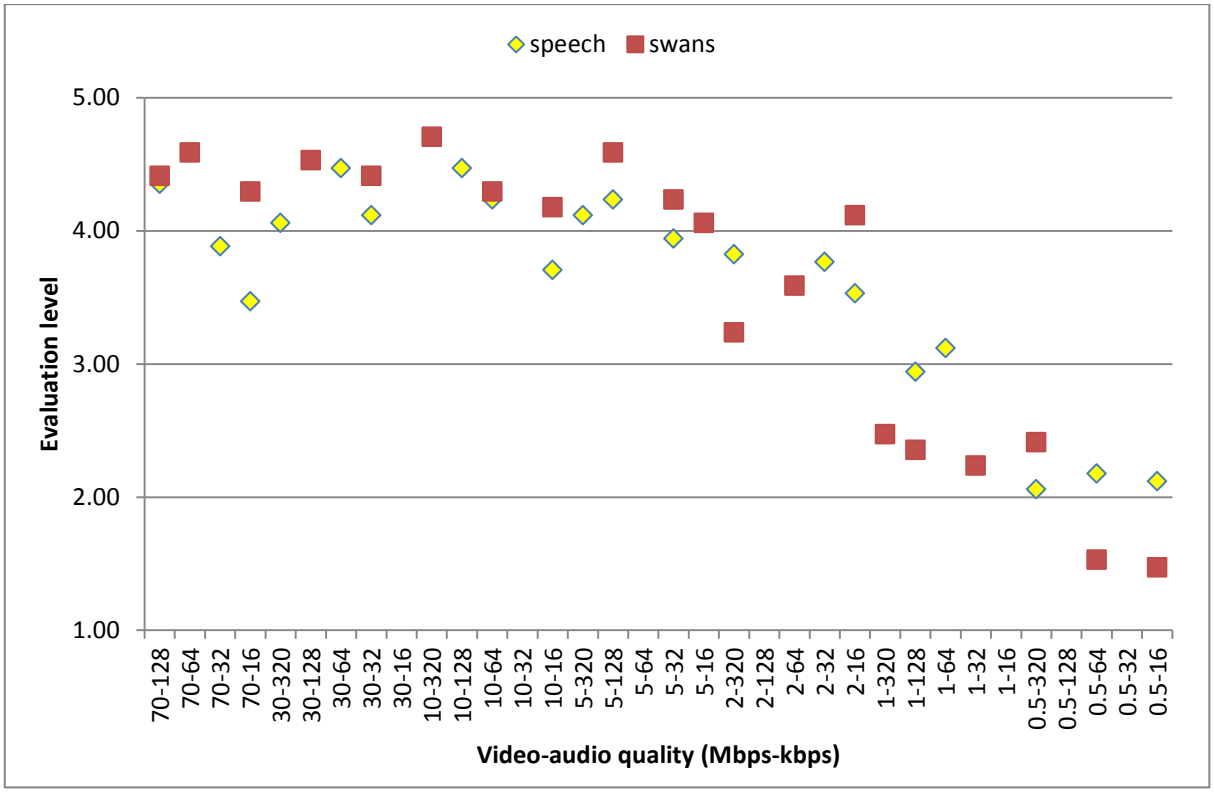


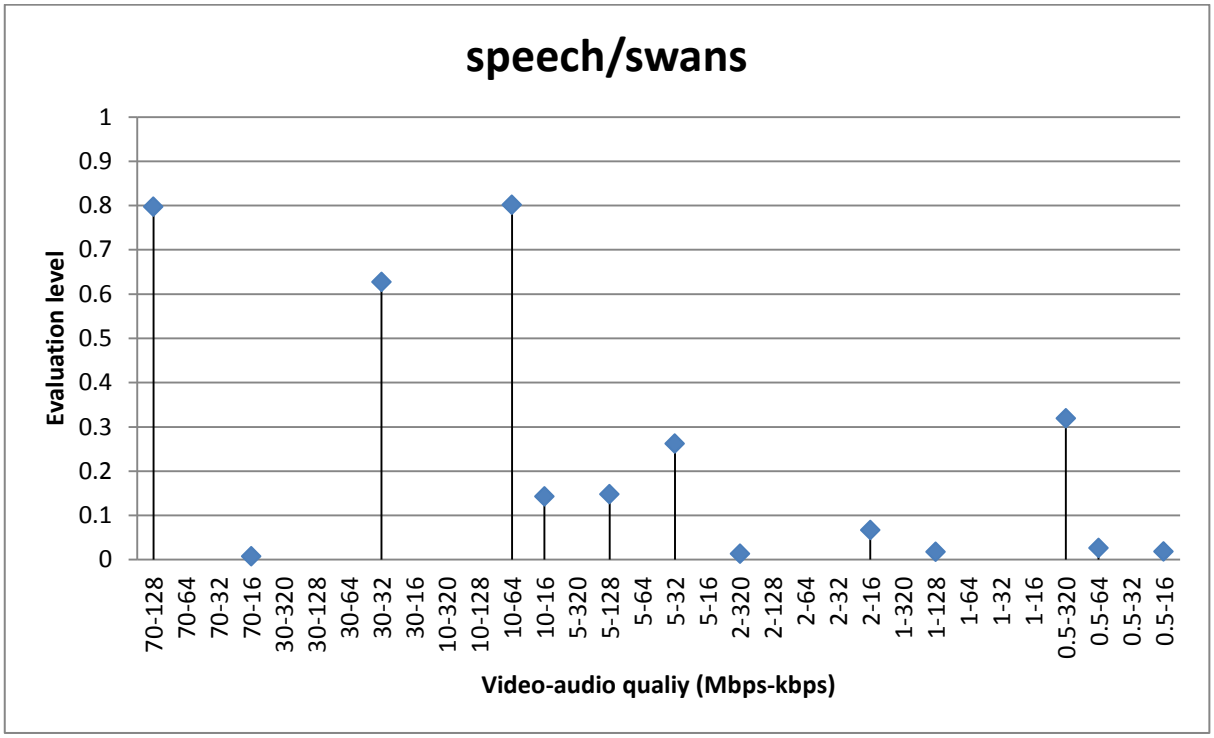
H) Speech tricks





I) Speech/swans





J) Speech/wall